THE UNIVERSITY *of* EDINBURGH

MAPPING GENE EXPRESSION TO FUNCTION IN ADULT MOUSE MEDIAL ENTORHINAL CORTEX

HELEN LUCY RAMSDEN

Doctor of Philosophy
School of Informatics
University of Edinburgh

2014

Helen Lucy Ramsden:
Mapping Gene Expression to Function in Adult Mouse Medial Entorhinal Cortex
Doctor of Philosophy, 2014

SUPERVISORS:
Dr. Matthew F. Nolan
Prof. J. Douglas Armstrong

LAY SUMMARY

The medial entorhinal cortex (MEC) is a region of the brain known to be involved in memory. Damage to the MEC or its neighbouring region, the hippocampus, disrupts learning and memory in humans and other animals, particularly when the memory has a spatial component, such as remembering the location of a parked car. Understanding the MEC is therefore an important focus of neuroscience research, particularly as it is one of the first regions to shown abnormalities in patients with Alzheimer's disease.

In this research my aim is to understand how gene expression can help us elucidate how cells in the MEC contribute to its function in the adult mouse, a commonly used animal model. Genes provide the instructions for the synthesis of proteins in cells. While every cell in the body contains a copy of its genome encoded in DNA molecules found in its nucleus, different cells and brain structures express different genes from the genome, which leads to the synthesis of different combinations of proteins. The consequence of this is that different cells in the brain have different functions. Thus, genes might contribute to the ability of the MEC to process memories, and could also contribute to the vulnerability of MEC to disease, so an important challenge is to identify them and the cells they influence.

To investigate how genes are expressed in MEC, I analysed data from the Allen Brain Atlas, an online database of experiments that reveal where in the brain the majority of genes in the mouse are expressed. I made several key findings. First, I identified over 2000 (10%) genes that are expressed at particularly high levels in a subset of MEC cells and that may provide a clue as to how cells function. Cells in the layer closest to the brain's surface, layer II, have previously been shown to represent space through patterns of electrical activity. The organisation of these cells is disrupted in Alzheimer's disease. Consistent with this, when I investigated the functions of the proteins encoded by genes with particularly high expression in this layer, I found that genes related to the cellular basis of learning, to parts of the cell that communicate with other cells, and to the molecular pathways disrupted in Alzheimer's disease are all highly likely to be expressed in layer II. Second, I discovered that over 1000 genes exhibit systematic differences in expression level in layer II cells positioned at different locations, a pattern that corresponds with differences in the spatial tuning of these cells. This group includes genes that perform the intriguing function of linking activity in neurons to learning, so their expression could help to understand how MEC cells learn about new environments.

To test whether an experience that stimulates activity in cells in MEC causes changes in the expression of one of these genes, I placed mice into a novel environment that they could explore. The expression level of the gene, which is essential for forms of learning, did not change in layer II, but increased significantly in a different cell population of MEC that receives inputs from the neighbouring hippocampus. It might therefore be predicted that learning due to a novel experience depends on interactions with the hippocampus. However, to test this, I

experimentally removed the hippocampal region that connects to MEC, and found that the expression of this gene continues to increase without the need for this input.

In summary, I have revealed that gene expression in the MEC is organised in a functionally relevant manner and I have identified genes that through highly specific expression may influence the ability of different MEC cells to perform certain tasks. Activity-dependent changes in gene expression can even provide insight into which MEC cells are involved in behaviours.

ABSTRACT

Deciphering the mechanisms that underlie circuit function in the hippocampal formation is a key challenge for neuroscience. This region, which includes the medial entorhinal cortex (MEC), is critical for spatial learning and episodic memory in humans. Spatially modulated cells in the MEC, the grid cells, provide a topographical representation of space, but we are yet to establish the neuronal properties that underlie this or the contribution that particular cells in different regions of the MEC and hippocampus make to circuit function. This is partially because the specific targeting of the network with genetic tools is complicated by a multitude of cell types with predominantly unknown molecular profiles.

To address our limited understanding of the molecular organisation of the MEC, I have characterised how the expression of genes is distributed throughout different layers of the MEC, using a custom-designed resource that facilitates analysis of *in situ hybridisation* data from the Allen Brain Atlas. Through simultaneous extraction of gene expression data across thousands of 2D aligned images, I reveal striking differences between layers within MEC, demonstrating that layer II contains the highest proportion of genes enriched in a single layer, whereas gene expression is very rarely confined to layer III. Of particular interest, layer II of MEC is highly enriched for Alzheimer's disease pathway genes, providing insight into its vulnerability as one of the first brain regions to show pathology. I also identify over 1000 genes that are expressed with a dorso-ventral gradient that maps onto the topographic organisation of MEC connectivity, grid cell spatial resolution and synaptic integrative properties of cells. An intriguing group of genes that closely relate circuit activity to gene expression, the plasticity-related activity-dependent genes, often show this pattern of expression.

Focussing on the activity-dependent expression of one such activity-regulated, plasticity-related gene, Arc, I provide a novel view of MEC function. During simple novel exploration, Arc expression is up-regulated to a much greater extent in the deep layers of dorsal MEC than in the grid cell-rich superficial layers. By selectively disrupting the predominant hippocampal input to dorsal MEC, which terminates in the deep layers, I show that the significance of this up-regulation is independent of hippocampal inputs. Thus, although research addressing MEC function is particularly focussed on the superficial layers, during the exploratory behaviour that potentially primes the system for representing an environment, important plasticity may be occurring at the synapses onto deep layer neurons.

In summary, my investigations of baseline and activity-dependent gene expression in MEC have revealed a molecular organisation both across different layers and along a functionally relevant gradient. This may be important for specifically targeting microcircuits in MEC and for characterising how laminar and regional differences contribute to the encoding of space in the hippocampal formation.

ACKNOWLEDGEMENTS

Four years ago I embarked on an exciting and challenging adventure that thanks to the help and support of a number of people has led to this thesis. First, I would like to thank the staff of the DTC: Jim Bednar, Mark van Rossum, David Willshaw and Peggy Series, for giving me the opportunity to undertake a PhD in neuroscience. I would like to thank my primary supervisor, Matt Nolan, for all the time he has invested in me and this project and for the teaching, guidance and advice he has given me, and Douglas Armstrong, my secondary supervisor, for his words of wisdom. I would like to thank my colleagues in the Neuroinformatics Doctoral Training Centre and in the Informatics Forum, for helping me with the technical side of my PhD and for many interesting discussions about the nature of neuroscience, the world, and how we should study these things. I would like to thank Steven McDonagh, in particular; without all his help with getting the image registration up and running on the ECDF cluster, the first two chapters might not be. I'd also like to thank Chris (Filo) Gorgolewski, Michael Hull and Chris Ball for introducing me to and helping me with Python and Numpy, and Ali Eslami for his insights on image segmentation. I am also indebted to members of the Nolan lab and the Duguid lab, who have committed time and effort to demonstrating experimental techniques and providing advice to me. In particular, I'd like to thank Christina McClure, Gülsen Sürmeli, Derek Garden, Melanie White, Arianna Rinaldi and Hugh Pastoll for their insights and advice. I would like to thank the animal facility technicians for all their help and positive attitudes on those long days spent there. I'd also like to thank all the staff who maintain the equipment I relied on, and Trudi Gillespie, for all her help with the confocal microscope.

A considerable part of this work has depended on data in the Allen Brain Atlas, so I would like to thank all those who have worked to produce this impressive resource. I'd also like to acknowledge all the mice that contributed to both parts of the project. From a distance, I must thank all the people who post Python and unix programming solutions to queries on Stack Overflow, and other such forums, and to those who have written plugins for ImageJ- these people all made the process of analysis far more fun and interesting than it might otherwise have been had I been limited to the simplest of tools.

Finally, I'd like to thank my parents, Anne and Chris, and wider family and friends for their support throughout this process.

DECLARATION

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

Edinburgh, 2014

Helen Lucy Ramsden, January 2014

CONTENTS

ABBREVIATIONS

MEC	Medial Entorhinal Cortex
ABA	Allen Brain Atlas
Arc	Activity-regulated cytoskeleton-associated protein
ROI	Region of interest
DV	Dorso-ventral
LTP	Long-term potentiation
LTD	Long-term depression
ISH	In situ hybridisation
GFP	Green Fluorescent Protein
SS	Somatosensory

1 INTRODUCTION

The medial entorhinal cortex (MEC) has been in the shadow of its neighbour, the hippocampus, since the discovery over 40 years ago that hippocampal neurons represent space (O'Keefe and Dostrovsky, 1971). Early on, it was predicted that the inputs to the hippocampus from the MEC might provide an important contribution, but it wasn't until the discovery of spatially modulated neurons in MEC (Fyhn et al., 2004; Hafting et al., 2005) that work began to address how the individual cell types and anatomical organisation of this system might support the interactions between spatial representations in the MEC and in the hippocampus.

The field is now at a point where MEC grid cells can be identified, recorded from and disrupted *in vivo* while an animal explores a real or virtual environment (Domnisoru et al., 2013; Bonnevie et al., 2013). However, the existence of spatially modulated cells across all layers of the MEC and the apparent broad distribution of inputs onto dendrites of cells across multiple layers has thus far prevented us from directly linking particular intrinsic features of cells to their circuit function in the behaving animal. Sophisticated tools that isolate circuit elements based on the genes they express could provide a solution, but they rely on an underlying knowledge of the molecular organisation of the system, which in MEC is relatively unexplored. In this thesis my aim is to provide a genetic description of the MEC that could both improve our understanding of laminar function and enable us to more selectively target components of the MEC circuit.

1.1 ORGANISATION OF THE THESIS

Chapter 2 provides a review of the literature spanning the three major topics covered in this thesis: the use of gene expression analysis as a neuroscience tool, the anatomy and function of the medial entorhinal cortex (MEC), and the role of activity-dependent gene expression in circuit function.

Chapters 3-7 are all results chapters that each contain their own introduction, methods, results and discussion. Chapters 3 and 4 cover analysis of gene expression data from the Allen Brain Atlas, whereas chapters 5-7 explore expression of the gene, *Arc*.

Chapter 3 is actually predominantly a detailed methods chapter, but ends with key findings. It outlines the use of the Allen Brain Atlas *in situ hybridisation* data and the development of a resource for its analysis. It also includes the primary, large-scale results of a gene expression analysis of MEC. Chapter 4 explores the functional significance of the results outlined in Chapter 3 and offers new insights into the organisation of MEC.

Chapter 5 predominantly describes the methods for analysing *Arc* expression. The results section presents an analysis of differences in *Arc* expression across regions at baseline, in preparation for Chapter 6. Chapter 6 describes the main results of the effect of novel exploration on *Arc* expression in MEC and other regions. This chapter reveals an interesting result

that is explored further in chapter 7. Chapter 7 describes methods for performing hippocampal lesions and outlines the effect of dorsal CA1 lesions on *Arc* expression in MEC.

The final chapter provides a discussion of results across all five results chapters and identifies the key findings of this thesis.

2

REVIEW

Each neuron within a circuit has a function to perform, a function that will be characterised by its molecular expression, morphology, synaptic connections, position within the network and its history of activity. In this thesis, one of my aims is to improve our understanding of the molecular profile of neurons in the medial entorhinal cortex (MEC). This literature review will explore both the research that has contributed to our understanding of the defining features of neuronal identity and function, and research that has investigated the organisation and function of neuronal circuits of the MEC. Finally, I will describe research into how activity-dependent genes can elucidate circuit function and relate this to the MEC and spatial navigation circuits. The questions outlined below have driven the choice of literature reviewed:

1. In what ways is a molecular understanding useful for understanding the function of a brain region?
2. Why is the MEC an interesting region for study?
3. What anatomical and physiological differences within MEC exist that lead to predictions about molecular expression?
4. In what way could theoretical predictions depend on or be modified by an improved understanding of molecules?
5. In what way do the interactions of the MEC with its surrounding brain structures drive predictions about the molecular signatures of its cells?
6. How can immediate early genes provide insights into neural circuit interactions?
7. In what way does immediate early gene-dependent plasticity contribute to spatial learning and the MEC?

2.1 WHAT DETERMINES NEURONAL IDENTITY AND FUNCTION?

Neurons in the adult brain contain predominantly identical DNA within their nuclei, yet their structure and morphology varies, they respond differently to inputs, exhibit different levels of activity and capacity for changing the strength of their synapses with other neurons, and as a result, their functions vary. This vast diversity confers our brains with highly complex and sophisticated circuits that we must isolate and dissect if we are to understand how the brain works.

Differences in neurons are driven by three major influences that interact with one another: the genes that a neuron expresses, its connectivity with other neurons, which determines what type of information its inputs carry, and its activity history. Here, in the first part of this review, I will introduce the parallel usage of genetics and connectivity to understand the nervous system and explain why investigating gene expression is a particularly useful focus at this time.

2.1.1 *The significance of genetic influences on neuronal function*

A major reason for differences amongst cells is that both throughout the development of an animal, and continuing into adulthood, the regions of DNA that are bound to and transcribed into RNA differ. In fact the patterns by which this occurs in adulthood will be strongly influenced by the processes that occurred during development. Although a complex series of events exists between transcription and the expression and function of certain proteins, this RNA can subsequently be considered as a template for the manufacture of the proteins that will influence the morphology, electrical properties and/or function of the neuron in a circuit. RNA expression levels may therefore allow us to identify potential combinatorial templates of genes that distinguish cells, and can be used in combination with protein expression to provide insight into a cell's function.

Cell-type specific gene expression: Gene expression alone does not usually indicate function because a neuron's role will be determined by its interaction with other neurons. However, many genes are expressed extremely selectively in particular neuron types (Sugino et al., 2006), often as a result of developmental processes, so analysing the functions of these genes can provide direct clues that relate genes to neuronal morphology, excitability, energy regulation, connectivity with other neurons expressing specific genes, and circuit function. A methodological advantage of profiling cell-type specific gene expression is that isolating such genes may allow cells to be selectively targeted. This could serve two purposes: (1) the manipulation of cellular activity in order to elucidate specific functions, and (2) identification of the full transcriptome and control mechanisms that regulate the expression of these genes, which could be useful for reproducing the behaviour in artificial or pathological circuits. Neuron-type specific targeting may also be useful for therapeutic interventions, since particular signalling pathways or molecules are sometimes disrupted only in specific cell types or structures; for example, the loss of dopamine in the substantia nigra in Parkinson's disease.

The tools available for measuring gene expression The detection of RNA has become an important neuroscience method and there are multiple approaches, some of which require explicit probes that bind to specific sequences of RNA, and others that identify all transcripts present through sequencing. *In situ hybridisation* (ISH) provides high cellular-level resolution of RNA level in cells. Because probes can be applied to tissue, *in situ*, transcripts can be precisely localised to brain regions (Lein et al., 2007) and even cell types (Kirsch et al., 2012). Although not specifically designed for quantifying the relative levels of different genes across different

locations [1], high-throughout efforts such as those by the Allen Institute to develop the Allen Brain Atlas (ABA), have provided high-quality data on brain-wide gene expression that is quantifiable within limits (Lein et al., 2007). This dataset can be used to extract differences in cortical expression patterns (Hawrylycz et al., 2011a, 2010, 2011b), locate genes with enriched expression at a specific coordinate in the brain (using the Anatomical Gene Expression Atlas), and locate genes with similar expression to other genes, either across the whole brain or in more confined regions using the NeuroBlast tool (Ng et al., 2007a, 2009). What is most valuable about this resource is that although expensive to develop, it can now be accessed and used flexibly by all, including for data mining purposes (Davis and Eddy, 2009) and, aside from maintenance of the website and data costs, is now free.

Microarrays provide increased quantifiability of RNA data because they directly and simultaneously compare gene expression. RNA-seq provides the additional advantage of using sequencing without the need for developing probes, and yields results with a very high dynamic range (Wang et al., 2009b). These methods often require pooling of tissue, thereby losing spatial resolution, but they can be used to analyse gene expression patterns from tissue containing only specific cell types, for example using fluorescence sorting, if a cell type has a known marker. The GENSAT project has developed a large number of transgenic mice that express Cre or a fluorescent reporter protein under the control of such markers (Heintz, 2004; Schmidt et al., 2013), but many cell types in the mammalian brain have not yet been isolated in this way. Gene expression therefore provides an open window into understanding cell function, but it often requires prior knowledge of defining molecular features to be extracted and used on a selective basis.

2.1.2 *The role of connectivity in determining neuronal function*

In the adult brain the majority of neurons exist in stable circuits with definable functions. A neuron's function is likely influenced by the function of the region in which it lies and the nature of the inputs it receives. The requirement to understand neural connectivity has driven the development of circuit diagrams, or connectomes; a connectome is not a universally defined entity, but incorporates a description of all the structural mappings and connections of all the neurons in the brain, the nature of the connection (excitatory or inhibitory) and the fine structural details, such as the number of vesicles at synapses (Lichtman and Sanes, 2008). The advantage of such a wiring diagram is that it could allow direct links to be made between structure and function, could elucidate circuit-level mechanisms and may enable us to understand why differences in brain organisation lead to differences in behaviour (Lichtman and Denk, 2011; Lichtman and Sanes, 2008).

The tools available: Mapping connectivity really began with the work of Cajal, who carefully detailed the precise morphology of Golgi stained neurons. His work dealing with structure enabled him to speculate about function, to the point of correctly identifying the direction in which electricity flows through neurons (Cajal, 1995). In the years since, scientists have developed a multitude of tools to probe projections (anterograde and retrograde tracers) (e.g.

1 Indirect methods must be used to detect binding and image analysis is required to extract the data

Dolorfo and Amaral, 1998b; van Groen et al., 2003), and actual synaptic connections (mainly using electron microscopy and heavy metal staining) (e.g. van Haeften et al., 2003; Schwerdtfeger et al., 1990).

Despite the potential benefits of creating a brain-wide description of structural connections, building a connectome poses serious challenges in mammals and some researchers believe its usage will be limited for understanding brain function. After all, the connectome of C-Elegans, which consists of just over 300 neurons, was identified over 25 years ago (White et al., 1986), yet our understanding of its behaviour remains limited. Mammals are even more complicated for several reasons, as summarised in Lichtman and Denk (2011): (1) the number of different cell types in the brain is vast, and means of distinguishing them are not well agreed on, (2) the extensive dendritic and axonal trees of neurons makes imaging the entirety of a single neuron and identifying its connections extremely difficult, (3) finding accurate and unambiguous tools for imaging electrical and chemical activity across many neurons, in order to establish functional relationships, is difficult, (4) there are limitations in the resolution of light microscopy that limit our ability to detect synapses, and (5) it is challenging to reconstruct all materials in the brain when detailed analysis is easiest with sparse labelling of neurons.

2.1.3 *The relationship between the connectome and gene expression*

Despite the difficulties discussed above, genetic approaches provide potential solutions to the challenges of elucidating the brain's structures. Gene expression characterises cell types (Sugino et al., 2006; Cahoy et al., 2008), and will aid in distinguishing different types of cells within circuits. Genetic manipulations can be used to express different coloured proteins across different neurons (e.g. the Brainbow mouse (Livet et al., 2007)), thereby enabling neighbouring neurons to be distinguished from one another. Deguchi et al. (2011) showed using mouse lines in which a fluorescent protein is expressed under the control of a modified Thy1 promoter that matched patterns of gene expression predict connectivity and neurogenesis time windows in the hippocampus, which emphasises how closely gene expression and wiring may interact. Finally, transsynaptic viruses, tagged with fluorescent proteins, provide a critical means of localising synapses using light microscopy (Lichtman and Denk, 2011), but rely on genetically defined cell-specific targeting.

2.1.4 *Why it is timely to study gene expression*

As discussed above, gene expression not only confers function, through the identification of proteins that are encoded in a cell, but can also provide a useful tool for understanding connectivity. While genetics and connectivity are often studied independently, there is clearly a large degree to which they can complement each other. Although there is unlikely to be a universal circuit organisation in any part of the mammalian brain, genetic identification of neurons, followed by selective manipulations, may allow us to identify both the circuit motifs that are generalisable across regions, and the elements that are most important. Recent advances in technologies that enable high-throughput RNA analysis techniques and protein-DNA in-

teraction research mean that neuroscience is ready for detailed transcriptomic analyses for cell types across the whole brain, but this first requires identifying molecular signatures for specific cell types. The relevance of large scale genomics projects has been reviewed by Luo et al. (2008), Geschwind and Konopka (2009) and Jones et al. (2009).

2.2 INVESTIGATIONS INTO THE GENETIC PROFILES OF CIRCUITS AND CELLS

2.2.1 *What do we know about global differences in gene expression?*

Thousands of genes show differential expression across the brain and they fall into particular functional categories. Over 6000 genes show patterned expression in mouse somatosensory (SS) cortex, according to RNA-seq data (Belgard et al., 2011) and high-throughput ISH data in the ABA (Lein et al., 2007), and many genes also show patterned expression in human neocortex (Zeng et al., 2012). Since many genes show conserved expression patterns between mice and humans (79%) (Zeng et al., 2012), it is likely that establishing patterns of gene expression in mice will be useful for understanding the importance of gene expression in humans. However, genes with specific laminar patterns are more likely than uniform genes to show different patterns between humans and mice (Zeng et al., 2012), which does somewhat limit the generalisability of findings.

Genes that are differentially expressed are often of functional interest in understanding what makes cell types and laminar functions different. Genes most likely to be expressed with a laminar or sparse pattern in human tissue include secreted proteins, peptide ligands and extracellular matrix proteins (Zeng et al., 2012), whereas synaptic proteins, types of nuclear protein and kinases are most likely to show widespread expression patterns (Zeng et al., 2012). In mouse SS cortex, genes involved in neurotransmitter receptor activity are highly likely to show patterned expression, whereas cell cycle and DNA binding genes are likely to be unpatterned (Belgard et al., 2011). One further advantage of identifying patterned genes is that disease genes are likely to show patterned expression (Belgard et al., 2011).

2.2.2 *The genetic dissection of individual cell types*

One of the contributions to patterned expression in the brain is that the majority of circuits possess a variety of intermingled cell types for which differences are not well defined. Even when they are well defined, perhaps on the basis of morphology or physiology, without a molecular description they are difficult to individually target, and so the intracellular mechanisms that underlie their physiology may not be well understood. Several projects have explored the molecular profile of cell types by exploiting resources that have developed transgenic mice that express fluorescent proteins under the control of specific promoters. The cells can be sorted using fluorescence-activated cell sorting (FACS) and the RNA extracted (e.g. Cahoy et al., 2008).

One of the first mammalian projects in which neuron types were molecularly characterised using transgenic lines investigated cell types under the control of the Thy1, Gad1 and Gad2 promoters (Sugino et al., 2006). These include several neocortical interneuron and projec-

tion neuron types, confirmed using electrophysiological profiling. Synaptic, axonal and membrane genes characterise individual cell types, along with actin binding, calcium-ion binding and ion channel activity-related genes (Sugino et al., 2006). Investigations into the classification of interneurons within the hippocampus have revealed that the calcium ion binding molecule that they express is closely related to morphology, electrophysiology and function (Toledo-Rodriguez et al., 2005).

In the retinal circuit different types of cells can be differentiated with reasonable confidence based on their positioning, morphology and physiology. In order to identify the transcriptomic code for these different cell types, Siegert et al. (2012) assembled a library of transgenic mouse lines that selectively mark a subset or single type of retinal cells using a fluorescent protein. They then used FACs to isolate specific cell types, and amplified the RNA in them to generate a retinal bar code. It transpires that each retinal cell type expresses a unique set of transcription factors and that cell transcriptomes, thresholded to only extract the most heterogenous genes, enable categorisation of cells into their specific groups. Siegert et al. (2012) were able to match cell expression with disease association, which could enable more specific treatments.

Cerebellar neurons can also be distinguished by their positioning and cytoarchitecture, a feature recently exploited to enable researchers to detect cell-specific molecules using classification of ISH images (Kirsch et al., 2012). They revealed the novel finding that a number of astrocyte-specific genes are located in the Purkinje cell layer (Kirsch et al., 2012). In the striatum, translational profiling has revealed that medium spiny neurons that are intermixed and appear highly similar can actually be differentiated based on their expression of very different protein-coding RNA templates (Doyle et al., 2008; Heiman et al., 2008). These differences are functionally relevant and highlight differential regulation of intracellular calcium (Doyle et al., 2008).

These studies suggest that a relationship exists between a neuron's molecular expression and its function, but there is also more direct evidence. The Drosophila circadian rhythm circuits consists of different functional groups of neurons, some of which are selectively labelled in transgenic flies in which gene expression is controlled under the UAS/Gal4 system (Nagoshi et al., 2010). By extracting RNA from specifically labelled cell types and performing a microarray analysis, Nagoshi et al. (2010) showed that a specific transcription factor, *Fer2*, is expressed in a particular functional group, which implies that its expression drives locomotor activity rhythms, while *nocturnin* mediates the circadian light response.

2.2.3 *Distinguishing neurons from astrocytes and glia*

Within circuits, neurons are accompanied by glial cells that include astrocytes and oligodendrocytes. Cahoy et al. (2008) sought to determine whether these cell types could be distinguished, again using a transgenic line, in combination with selective methods to extract or deplete oligodendrocytes (OLs) using immunohistochemistry and immunopanning, respectively. Neurons, astrocytes and oligodendrocytes exhibit distinct markers and their profiles show similar levels of variation from one another, suggesting that astrocytes and OLs, although both termed glial cells, are likely to be as molecularly distinct as neurons and as-

trocytes (Cahoy et al., 2008). Functionally, these different cell types are enriched for genes involved in a variety of signalling pathways. While genes in all three groups are enriched for axonal guidance signalling, calcium signalling is only enriched for neuronal genes (Cahoy et al., 2008).

2.2.4 *Gradients in gene expression*

I have previously discussed the cell-type expression of genes as though populations of cells express genes uniformly. Classically in development, but also in adulthood, gradients in gene expression across space or across a population drive differences in function. At the optic chiasm, molecular gradients guide axons to the contralateral side during development. The precision with which activity guides axons to form contacts within the optic tectum of the Xenopus is driven by an opposing gradient in Wnts and EphBs in the ventral and dorsal tectum, respectively (Lim et al., 2010).

In adults, several molecular gradients have been observed throughout the hippocampus and surrounding regions. Thompson et al. (2008) used the ABA data to assess changes in gene expression throughout the septo-temporal extent of the hippocampus. They showed, using a tool to segregate anatomical regions based on degree of similarity [2], that CA3 can be genetically segregated into 9 components based on gene expression, in particular the expression of transcription factors, ion channels and adhesion molecules (Thompson et al., 2008). Similar gradients have been observed in CA1, where pyramidal cells can be subdivided based on characteristic laminar distribution of certain molecules (Dong et al., 2009). In the medial entorhinal cortex, which projects topographically to the hippocampus, several molecular gradients have been reported using data from the Allen Brain Atlas (Garden et al., 2008). As will be discussed in Section 2.3.3, these molecular differences map onto functional differences in the spatial modulation of cells (Hafting et al., 2005), potentially through tuning of synaptic responses (Garden et al., 2008; Narayanan and Johnston, 2008; O'Donnell and Nolan, 2011).

2.2.5 *The regulation of gene expression*

The reason that cells can be defined based on differential gene expression is because certain transcription factors and regulatory RNAs maintain differences in the sequences of code that are actively bound and ultimately code for proteins. The roles of the functional elements of DNA have received similar experimental and theoretical (Hobert et al., 2010) attention to the genes themselves, as identifying them could provide us with the raw code that defines a cell. They could be particularly useful for reprogramming stem cells to become particular cells (Caiazzo et al., 2011).

Large projects are underway to identify all functional elements of the human genome (Consortium, 2012), the mouse genome, which shares a high degree of conservation with the human genome (Mouse Genome Sequencing Consortium et al., 2002), and Drosophila genome (modENCODE Consortium et al., 2010). Currently, chIP-Seq methods, which identify cis-

2 Non-negative matrix factorisation was used on 200μm voxel data from the hippocampal region, using 3D neighbouring constraints

regulatory elements by extracting sections of DNA that are bound to proteins, provide the most effective way of revealing a code. Recent chIP-seq work has allowed the mapping of nearly 300,000 cis-regulatory sequences, which accounts for 11% of the mouse genome (Shen et al., 2012). By using many different tissues, Shen et al. (2012) were able to identify tissue-specific enhancers, including for the mouse cortex, and identify potential TFs that regulate expression in each type.

It should be noted here that the function of a cell is predominantly affected not by the transcribed genes but by the proteins expressed. Protein levels depend not only on mRNA expression, but also RNA degradation and trafficking, translational mechanisms and post-translational modifications. While this is an important feature to keep in mind, sufficient analysis of the functional elements of DNA may enable estimation of the significance of the transcriptome to cell function across different neuronal types. Periods of plasticity are one of the key times when both gene and protein expression in cells will be highly dynamically regulated. The immediate early genes, which will be discussed in more detail in Section 2.4, are well known as indicators of neuronal activity because their RNA can be transcribed within minutes of a stimulus. The expression levels of many other genes are also altered following memory tasks (Rapanelli et al., 2010; Klur et al., 2009; Miyashita et al., 2009). Intriguingly, differences in the expression patterns of genes across hemispheres following spatial tasks has indicated lateralisation of hippocampal function (Klur et al., 2009).

2.2.6 *Gene expression and disease vulnerability*

The molecular identity of cells is not only important for normal function, but provides strong indications that cell types may be susceptible to abnormal function, since mutations and variants in specific genes have been associated with disease. Ion channels and synaptic genes, particularly those in the post-synaptic density, play an important role in the activity levels and computations of a cell, and are often closely associated with disease (Grant, 2005).

2.2.6.1 *Genes linked to neurodegenerative disease*

The neurodegenerative disease Alzheimer's disease (AD) affects 1 in 6 people over the age of 80 (Alzheimer's Society). It is characterised behaviourally by a progressive decline in memory and other cognitive functions (Huang and Mucke, 2012). It is named after Alois Alzheimer, a neuropathologist who first described the amyloid plaques and neurofibrillary tangles that are now considered hallmarks of the disease. Although these features are common to all forms of AD, the disease has multiple causes (Huang and Mucke, 2012). Most notably, although it is estimated to be highly heritable, different genes have been associated with the disease, and only in the case of familial, early-onset AD ($\sim$ 1% of all cases), is it caused by autosomal dominant mutations (summarised in (Bertram and Tanzi, 2008; Huang and Mucke, 2012)). Over 500 candidate genes have been associated with late-onset or sporadic AD, but few have been replicated sufficiently to suggest that they are consistently involved (summarised in (Bertram and Tanzi, 2008)).

Amyloid plaques appear to result from dysregulation of the amyloid-β pathway (summarised in (Huang and Mucke, 2012)). Mutations in the genes that cause familial AD, APP,

Psen1 and Psen2 disrupt this pathway (summarised in Bertram and Tanzi (2008)). While Aβ is generated from amyloid precursor protein (APP), Psen1 and 2 are genes that encode proteins that affect the sequential cleavage necessary for this. However, amyloid plaques do not correlate well with cognitive decline (Braak and Braak, 1991). Candidate gene and genome-wide association studies have consistently identified the apolipoprotein E (ApoE) locus as an important determinant of AD risk and age of onset (Strittmatter et al., 1993; Farrer et al., 1997). ApoE both directly and indirectly influences Aβ production. The variant associated with increased risk of late-onset AD, E4, impairs Aβ clearance and promotes deposition but is also activated by neurons under stress, causing neurotoxic fragments that disrupt other cell functions. Systematic meta-analyses have shown that around 20 other candidate genes, including ABCA7, BIN1, CD2AP, CD33, CLU, CR1, EPHA1, MS4A, and PICALM may have modest but significant effects on risk of AD (summarised in (Huang and Mucke, 2012)).

Despite advances in our understanding of the genetics of AD, we do not understand the full aetiology and remain unable to treat the neurodegeneration or symptoms effectively. The reasons for this include that AD genetic risks likely interact with environmental features, including hypertension and obesity, animal models are often limited to one aspect of the genetic disruption or pathology, and the pathological features that correlate most clearly with cognitive decline are not necessarily those directly affected by mutations. These include aberrant neural network activity, dysfunction and loss of synapses and degeneration of specific neuronal populations, which are likely to be downstream effects of Aβ disruption (Huang and Mucke, 2012). The cause of the tau dysregulation that leads to neurofibrillary tangles is also unclear. Other downstream effects include changes in receptor distribution, intracellular calcium homeostasis, interneuron activity, axonal transport and mitochondrial function (Huang and Mucke, 2012). If we are to understand what makes certain brain regions and cells vulnerable to degeneration in AD, it may be important to consider not only whether mutated genes are strongly expressed in them, but whether genes throughout the affected pathways are highly expressed.

Other neurodegenerative diseases include Parkinson's disease and Huntington's disease, which predominantly disrupt motor control and may also affect cognition. Mutations in LRRK2, PARK2, PARK7, PINK1, or SNCA are the genes most robustly associated with Parkinson's disease (PD), where mutations in SNCA and LRRK2 can cause autosomal dominant PD (reviewed in Lesage and Brice (2009)). Most recently, axon guidance molecules such as DCC, EPHB1 and NTNG1 have also been implicated in PD (Lin et al., 2009). Huntington's disease is caused by an autosomal dominant mutation in the huntingtin (Htt) gene and its severity and age of onset depends on the number of CAG repeats in the Htt gene.

Although symptomatically different, what unites Alzheimer's, Parkinson's and Huntington's disease is that they involve progressive but extensive cell disruption and loss in later life. They also consistently cause abnormalities in particular brain regions, including the parahippocampal formation, which suggests that these regions may be similarly vulnerable to the dysregulated pathways that lead to degeneration. The Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa et al., 2004) pathways of these diseases contain ~ 130 genes, many of which are shared. These include mitochondrial genes, calcium-signalling-related genes and apoptosis-related genes.

2.2.6.2 *Genes linked to neurodevelopmental disorders*

The second type of disorder with genetic links are neurodevelopmental disorders, which include schizophrenia, epilepsy and autism. Epilepsy, which is characterised by sporadic uncontrollable seizures, has been associated in its familial form with mutations in ion channels, specifically sodium channels (Catterall et al., 2008), which leads to reduced ability of circuits to control neuronal excitability. However, studies examining the genetic basis of the sporadic form of the disease suggest that extremely complex networks of mutations likely occur to cause pathology (Klassen et al., 2011) and that non-ion channel genes and pathways are also involved (Kasperaviciute et al., 2010). Sophisticated computational models are likely to be necessary to understand how mutations, dysregulated pathways and disease interact (Kim et al., 2011).

Schizophrenia is a disorder characterised by the 'positive' symptoms, hallucinations and delusions, and the 'negative' symptoms of lack of affect and has been linked to abnormalities in neuronal connectivity (Goff and Coyle, 2001). Schizophrenia affects approximately 1% of the UK population and a large number of genetic regions of interest have been implicated through association, linkage and assessment of submicroscopic copy number variations studies. The most influential candidates are spine-enriched genes Neuregulin-1, ErbB4, RGS4, CAPON, COMT and DISC1 (summarised by Harrison and Weinberger (2005)). Autism is also a highly complex genetically-associated disorder, with few robust candidate genes. Recent emphasis has been placed on a neurexin family gene, CNTNAP2, based on results of association studies and on its enrichment in developing circuits involved in language processing (Alarcón et al., 2008). This gene may also be implicated in schizophrenia pathology, thereby linking the disorders (Burbach and van der Zwaag, 2009).

2.3 MEC AS A MODEL SYSTEM FOR INVESTIGATING GENE EXPRESSION

The medial entorhinal cortex (MEC) is a brain structure that, since the time of Cajal (as cited in (Canto et al., 2008)) has received considerable interest because of its positioning as the gateway between the hippocampal formation, a limbic structure with particularly interesting architecture, and the neocortex. The MEC is an attractive structure for considering gene expression for several reasons. First, its general function as a component of the spatial navigation system is well established, of great interest, and appears to show many parallels in lab animals and in humans. Specifically, it contains a type of cell that represents spatial location, the ' grid' cell (Hafting et al., 2005). Second, its structure is unlike the structure of other cortical regions. Although it possesses a six-layer cortex, the organisation of layers and cell types differ from neocortex, which may be related to its molecular organisation. Third, the MEC shows a gradient in the physiological and morphological properties of cells that directly maps on to the resolution with which its cells encode space (Hafting et al., 2005; Garden et al., 2008; Giocomo and Hasselmo, 2008). Fourth, despite considerable interest and considerable research, we lack a clear characterisation of the different neuronal subtypes, we cannot genetically isolate particular cell types, which dramatically limits our capacity to understand the full function of the circuit, and we do not know where the learning that drives changes in the representation of space occurs in the circuit. Finally, the MEC appears to be particularly vulnerable to pathology associated with a number of different diseases including Alzheimer's disease (Braak and Braak, 1991), schizophrenia (Arnold et al., 1991), epilepsy (Du et al., 1993) and autism (Bauman and Kemper, 1985). In this section of the review I'll outline research into the function, anatomy, connectivity, and disease associations of the MEC.

2.3.1 *The MEC as a component of the spatial navigation circuit*

From rodents to humans, the MEC is a well-recognised component of the spatial navigation circuit. The hippocampus and entorhinal cortex interact with other neocortical and allocortical regions to provide a geometric (based on landmarks) and idiothetic representation (using path integration) representation of space (for reviews, see (Witter and Moser, 2006; Moser et al., 2008)). To do this requires processing of location, direction and distance information from a plethora of brain regions and the subsequent analysis of angular and linear self-motion (for reviews, see (Witter and Moser, 2006; van Strien et al., 2009; Jeffery, 2007)). However, the precise roles played by different regions remain unclear. The function of the lateral EC, which also projects to the hippocampus, is not well established, but its cells are only weakly spatially modulated in behaving animals (Hargreaves et al., 2005) and do not represent location unless discrete objects are present (Tsao et al., 2013). In this review its connectivity and function will not be discussed.

2.3.1.1 *The hippocampal-entorhinal circuit*

The hippocampus and its posited function: The hippocampus was convincingly implicated in memory following the case of H.M., who suffered retrograde and anterograde amnesia following bilateral surgical ablation of his hippocampus, and other regions (Scoville and Milner,

1957). At this time, it could not be ascertained how much individual memory processes, such as acquisition, storage, consolidation and retrieval of memory depended solely on the hippocampus, but more recent work has shed light on this. The hippocampus has been shown to play clear roles in the acquisition and consolidation of spatial reference memory (Morris et al., 1982; Moser et al., 1993), spatial working memory (Niewoehner et al., 2007) context-dependent memory when a delay is involved (Ainge et al., 2007), salient events in temporal sequence that involve nonspatial stimuli such as texture, and odours (Wood et al., 1999), and episodic memory in humans (Scoville and Milner, 1957). In all these cases the hippocampus appears to be particularly critical for acquisition and consolidation, while cortical regions may be more heavily involved in storage and retrieval processes, particularly for very long-term memory (van Strien et al., 2009). Although the hippocampus has been implicated in navigational processes, object-place associations, and fear memories, there is considerable evidence that different processing is involved (Bachevalier and Nemanic, 2007). For a review of the involvement of the hippocampus in memory, see (van Strien et al., 2009; Bird and Burgess, 2008; McNaughton et al., 2006a; Moser et al., 2008). The specificity of the hippocampal role in certain tasks will also be discussed in Chapter 6 7.1.

The hippocampus does not contribute to these processes in its entirety; instead its subregions function in specific aspects that depend on the precise task requirement. In 1971 an important discovery was made that shed light on the involvement of the hippocampus in spatial memory. In the CA1 subfield on the hippocampus O'Keefe and Dostrovsky (1971) discovered spatially modulated cells that fire only when an animal is in a particular location in its environment. Place fields have also been reported in cells in the CA3 subfield and dentate gyrus (DG) (O'Keefe and Dostrovsky, 1971), as well as subiculum, parasubiculum and presubiculum (Taube, 1995; Sharp and Green, 1994), although CA1 place fields remain the most well-studied. However, we do not yet have a complete understanding of how place fields develop or how closely they relate to spatial memory.

Place cells in CA1 show increased stability as an animal matures, they persist in the absence of intrahippocampal input from DG (McNaughton et al., 1989) and CA3 (Brun et al., 2002)[3] and they can be reset by changes in landmark position (Gothard et al., 1996a). It has been suggested that space and temporal sequences can both be represented by place cells, but using a different code; while space may be represented by the locations at which a cell fires, the relative timing of cues may be represented by cell firing rate(Leutgeb et al., 2005). Evidence from long-term calcium imaging studies suggests the same place can be represented by multiple cell ensembles, where different combinations are recruited at different times(Ziv et al., 2013). Of relevance to the MEC-hippocampal circuit, place field stability and selectivity depends on input to CA1 from the MEC (Brun et al., 2008a; Steffenach et al., 2005) and it is NMDA-dependent (Kentros et al., 1998). However, the formation and short-term stability of place fields do not depend on NMDA activation (Kentros et al., 1998).

Connectivity: The extensive connectivity between MEC and the hippocampus provides strong evidence that the MEC might have similar or complementary functions to the hippocampus. Evidence suggests that all regions of the MEC project to all regions of the hip-

3 although this research did find selective disruption to spatial learning and recall

pocampus, and vice versa (summarised in (Kjonigsen et al., 2011; Canto et al., 2008)), but some of these projections dominate. One of the features of this circuit that makes projection targets difficult to ascertain is that dendrites from cells in all layers are present in the superficial layers (summarised in Canto et al. (2008)). A second feature is that long-range projections between MEC and hippocampus have been shown to be both excitatory and inhibitory (Melzer et al., 2012).

The MEC provides the major input that the hippocampus receives, via the perforant path (PP) and temporo-ammonic (TA) pathway. While layer II axons target dendrites in the DG and CA3 via the PP (Steward and Scoville, 1976; Dolorfo and Amaral, 1998b), layer III axons extend via the TA pathway to CA1. The CA2 region of hippocampus also receives inputs from MEC onto its distal dendrites that originate from layer II and III neurons and that show similar timing to DG inputs from the PP (Bartesaghi and Gessi, 2004; Chevaleyre and Siegelbaum, 2010). PP inputs from the MEC to the hippocampus selectively innervate particular layers of the DG, CA2 and CA3, thereby targeting particular regions of the pyramidal or granule cell dendrites along the entire transverse axis (Witter et al., 1989). The TA input targets the dendrites of the proximal half of the transverse axis of the CA1 region, which is close to CA2 and distant from the subiculum (Witter, 1993). Inputs from the LEC innervate the converse regions, a finding that led to predictions that proximal CA1 cells would be more spatially tuned than distal CA1 cells. This has since been confirmed in experiments investigating differences along the transverse axis in extracellular firing patterns (Henriksen et al., 2010), and in immediate early gene expression following a spatial task (Hartzell et al., 2013). Anterograde and retrograde tracer studies have shown that the major output from the hippocampus to the MEC is from proximal CA1 to the deep layers, in particular the inner layer V (Swanson and Cowan, 1977; Tamamaki and Nojyo, 1995). These layers also receive input from subicular cells that receive their prominent input from CA1 (Tamamaki and Nojyo, 1995; Kloosterman et al., 2003). The major limitation of tracing experiments is the inability to precisely localise the site of the injection to specific neuron types, and difficulty in identifying the neurons targeted by labelled axons. These difficulties can be overcome with the use of transgenic mice that can selectively express a retrogradely transportable virus in select cell types, given that the cells can be molecularly targeted (Weible et al., 2010). To quantify monosynaptic inputs to layer II cells in MEC, Rowland et al. (2011) used this technique and revealed a previously unreported but substantial projection from area CA2, which itself receives a projection from layer II and III neurons that is capable of plasticity (Chevaleyre and Siegelbaum, 2010).

These patterns of connectivity, which are highly topographic, may well provide important clues as to the different functions of both the hippocampal subregions and the MEC layers. In particular, the input/output distinctions across the layers may have important implications for different roles in memory. A summary of the circuit is shown in Fig. 2.1(A).

2.3.1.2 *Evidence for the role of MEC in spatial memory*

The role of the hippocampus in spatial memory, combined with it extensive connectivity with MEC, have led to both theoretical insights and experimental research into the specific role played by the MEC in memory. Almost 40 years ago, O'Keefe (1976) posited that the place cells he had discovered were likely to receive important inputs from outside of the hip-

pocampus. Since then, many phenomenological and computational theories have attempted to explain the relationship between MEC input and the formation of place cells. Many make very specific assumptions about the architecture of the intraconnectivity and the intrinsic properties of the cells. For some, MEC is thought to hold a universal memory representation, which the hippocampus uses to build up context specific memories (for reviews see (van Strien et al., 2009; Moser et al., 2008)). However, studying the roles of MEC and hippocampus independently has proved difficult, and due to the extensive connectivity, manipulations to either may have considerable feedback effects. Nevertheless, there is now convincing evidence that the MEC plays a specific role in spatial memory that is beyond simply providing a relay input.

Two experiments have clearly demonstrated that the TA pathway from MEC layer III to CA1 is critical for spatial processing. Specific lesions to the TA pathway prevent the consolidation, but not the acquisition of short-term memory or the retrieval, of long-term memories for the location of a platform in the Morris water maze task (Remondes and Schuman, 2004). One interpretation of the data is that although this pathway may be important for precise spatial memory, there is evidence that TA-lesioned animals can acquire and retrieve a less accurate form of long-term memory, presumably using a different pathway (Remondes and Schuman, 2004, see Fig 2d). An interesting question is how flexible this circuitry is; could animals given more extensive experience in spatial tasks following an adult lesion eventually perform just as well on the task by learning to use an alternative strategy that recruits a different entorhinal-hippocampal pathway. The second experiment explored place cell stability and showed that specific lesions to layer III of MEC disrupted CA1 place field stability, but not CA3 place field stability (Brun et al., 2008a). This research may therefore indicate different forms and requirements for information processing in the two superficial layers of MEC.

Other research has shown that neonatal lesions to MEC a week after birth disrupt spatial working memory in the 8arm radial maze (Harich et al., 2008). The radial maze can be used for working memory tasks because it is possible to completely dissociate spatial reference and spatial working memory by baiting multiple arms and testing whether animals visit the same arm twice (see Niewoehner et al., 2007). This experiment therefore shows that MEC circuitry must develop normally for intact spatial working memory, but does not provide insight into the role of MEC in spatial working memory in adulthood. Experimental work identifying neural signatures for working memory placed emphasis on the MEC as an important structure because of the particular properties of layer V cells (Egorov et al., 2002). However, the difficulty in isolating specific features of MEC-entorhinal circuitry, as well as the involvement of prefrontal circuits, has left a specific layer V role unclear. A recent study suggests that in humans, stimulation of the MEC, but not the hippocampus, may improve performance on space-related tasks (Suthana et al., 2012). Together, these experiments provide strong evidence for a role of the MEC in tuning the properties of place fields and in interacting with the hippocampus to process spatial memory, but do not provide insight into a mechanism, beyond the possible need for plasticity in the TA pathway synapse during memory consolidation.

2.3.1.3 *Grid cells in the MEC*

The discovery of spatially modulated cells in MEC (Fyhn et al., 2004) has driven more clearly defined mechanistic explanations for the interactions between MEC and hippocampus that underlie spatial memory, and has increased the focus on the organisation of the MEC. These cells, now termed 'grid' cells (Hafting et al., 2005), have multiple firing fields that extend across an environment and are aligned in a regular hexagonal pattern (Hafting et al., 2005). They are found across all layers of MEC and throughout the dorso-ventral and medio-lateral extent, although topographical variations do exist (Fyhn et al., 2004; Hafting et al., 2005). Most salient, the spacing and size of grid fields increases along the dorso-ventral axis (Hafting et al., 2005), but not along a continuum. Rather, discrete modules of cells exist that exhibit similar grid firing characteristics (Stensola et al., 2012). In dorsal regions, just one or two modules operates, whereas more ventrally, a larger number of modules are active, and the distribution of modules is similarly spread across the controlateral MEC (Stensola et al., 2012). This has important implications for the organisation of cells within MEC, as well as for the connections that maintain similar behaviour across different hemispheres.

Grid cells are present not only in rats, but in mice (Fyhn et al., 2008) and bats (Yartsev et al., 2011), and there is evidence from fMRI imaging that grid cells also exist in humans (Doeller et al., 2010). Grid cells are accompanied by head-direction cells (Sargolini et al., 2006), border cells (Solstad et al., 2008), and conjunctive head-direction and grid cells that are all responsive to movement (Fyhn et al., 2007). The behaviour of grid cells has been described as one of the most conspicuous behavioural correlates that exists beyond that of sensory neurons (Moser et al., 2008). As such, understanding the mechanisms that underlie grid like behaviour, both intrinsic, which may be genetically regulated, and activity- and connectivity-dependent, is a highly active research area (for a review, see Moser et al. (2008)).

Grid cells appear from the first entry made into a novel environment (Hafting et al., 2005). The grids of neighbouring cells are offset relative to each other, such that phase is randomly organised across the MEC (Fyhn et al., 2007). Grid field spacing and orientation preference, by contrast, are topographically organised (Hafting et al., 2005), suggesting that the organisation of inputs and properties controlling these features is differentially organised. The properties of neighbouring cells remain constant across different environments (Hafting et al., 2005), which suggests that intrinsic properties and hard-wired connectivity play an important role in grid cell organisation (Jeffery, 2007). It remains unclear whether molecular differences in cells predict such organisation. Grid cells persist after the removal or replacement of major landmarks (Hafting et al., 2005; Sargolini et al., 2006), suggesting that they are driven and updated by self-motion information. Such inputs are likely to be received by cells in multiple MEC layers (Canto et al., 2008), but the precise identity of the receiving cells has not been elucidated. Experience can modify grid properties; grid fields scale with changes in environment size and are affected predominantly by extrinsic cues (Barry et al., 2007), and they fragment in environments with internal divisions (Derdikman et al., 2009). The most convincing evidence that grid cells in MEC interact with hippocampal processing is that grid fields are disrupted by muscimol-controlled inactivation of the hippocampus (Bonnevie et al., 2013). Further to this, global remapping in place cells is accompanied by realignment and remapping of the grid map (Fyhn et al., 2007), suggesting that they directly interact. The MEC is therefore not

only ideally situated and connected for spatial processing, but also possesses cells that metrically encode space and are organised in a manner that emphasises the importance of local differences in properties.

2.3.2 *The structure and cell types of the MEC*

The previous section highlights the functions that have been attributed to the MEC-entorhinal system. As yet, very little is known about which components of the system contribute to specific functional features and whether molecules, anatomy and connectivity predict observed differences. In this section I will outline the known anatomy and physiology of MEC and its cell types. What will emerge is that there is a vast amount of interconnectivity and topography within the entorhinal-hippocampal loop, suggesting many parallel and complementary routes of information processing. However, robust patterns are rarely observed due to limitations in our ability to classify groups of neurons and identify their specific function.

2.3.2.1 *The MEC as a whole*

The MEC possesses a large range of cell types that are positioned across its 6 layers with different degrees of dendritic arborisation and axon branching (Canto et al., 2008). Despite its apparent 6-layer organisation, some have argued that its structure falls between that of 3-layered archicortex, such as found in parasubicular areas, and the neocortex (Stephan, as cited in Canto et al. (2008)). In general, all MEC layers project to the hippocampus but the precise location to which they project depends on cell type and layer (see (Kjonigsen et al., 2011) for an atlas of connections). This complex organisation makes establishing a molecular profile all the more important; if inputs are targeted to particular molecularly defined cells, a molecular description will aid in distinguishing them and even if inputs are not specifically targeted, a molecular profile will provide insight into how cells could process the inputs differently.

Molecular markers that distinguish the MEC from some of its neighbouring structures are already known and are important complements to cytoarchitectonic and projection-based division schemes. MEC can be distinguished at the lateral extent from perirhinal cortex and at the dorsal extent from postrhinal regions by the presence of parvalbumin-positive neuropil, low calbindin labelling and low Timm's staining (summarised in Canto et al. (2008)). Corresponding cytoarchitectural markers of MEC include the large sized cells of MEC, which are present in layer II and disappear at the dorsal border, and the presence of a laminar dissecans (Insausti et al., 1997). The medial border with the parasubiculum is currently more difficult to use molecular markers for, but is typically marked by the loss of differentiation between the deep layers at the medial border (Insausti et al., 1997).

There are no known molecular markers that clearly distinguish the internal lateral and medial EC division. Brodman divided the EC into two regions based on cytoarchitecture, regions that approximately correspond to a medial and lateral EC (as cited in Canto et al. (2008)). Lorente de No (as cited in Canto et al. (2008)), using connectivity profiles, distinguished a third region at the ventral extent, the ventromedial EC. However, the projections to CA1 do not clearly delineate these boundaries, so more recent divisions have used the fact

that MEC neurons project to the outer layer of the dentate gyrus and receive inputs from the presubiculum (although only in the dorsal region) whereas LEC does not receive presubicular input (Witter et al., 1989)

Intrinsic Organisation One of the intriguing features of the MEC that is also found in the hippocampal circuitry is the extensive intrinsic, recurrent connectivity (Köhler et al., 1986) that exists within and across all layers. It may be of functional significance as the firing characteristics of grid cells are correlated across layers when a rat performs a spatial task (Sargolini et al., 2006). Recent work has placed particular emphasis on the modular organisation of the intrinsic connectivity (Witter and Moser, 2006), and has drawn attention to periodic changes in the expression of molecular markers that map onto changes in neurons, intrinsic circuitry, and dendritic and axonal clusters (Witter and Moser, 2006). Layer II cells appear to be organised into patchy islands while discrete bundles of layers III and V pyramidal dendrites (Ikeda et al., 1989) as well as deep layer axons (van Haeften et al., 2003) extend throughout the deep and superficial layers. These bundles show variability in the densities of GABA positive and calbindin positive dendrites (Solodkin and Van Hoesen, 1996; Suzuki and Porteros, 2002). The density of the presubicular inputs that define the extent of MEC also show cyclic variations (Caballero-Bleda and Witter, 1993). These data suggest that modules exist within the organisation of MEC. If so, maintaining the specificity of connections within and between such modules could rely on molecular markers that act as axon guidance molecules and adhesion molecules.

Inputs and projections of the MEC: The MEC, as discussed previously, is situated as a gateway between the hippocampus and neocortex. It also receives inputs from and sends projections to the basal forebrain, claustrum, amygdala (Pikkarainen et al., 1999), basal ganglia, thalamus, hypothalamus, parahippocampal region, and brainstem (reviewed by Burwell and Witter (2002)). A subset of these inputs and outputs are shown in Fig. 2.1(B). The function of some of these inputs is known, while others are not. For example, head-direction modulated inputs that likely tune the head-direction preferences of MEC cells originate from the presubiculum (all layers) (Canto et al., 2012) and anterior thalamus (into layer V), parasubiculum (all layers) (Canto et al., 2012), and the RSC (predominantly to layer V) (Jones and Witter, 2007). These structures are reviewed in (Taube, 2007). Any functional study of the MEC should take the effects on these circuits into account, but the identity of the neurons involved in these circuits is not well known, which makes it difficult to control. In the next section I will discuss what is known about the anatomy of the different layers of MEC, individually, as well as the cell types within them.

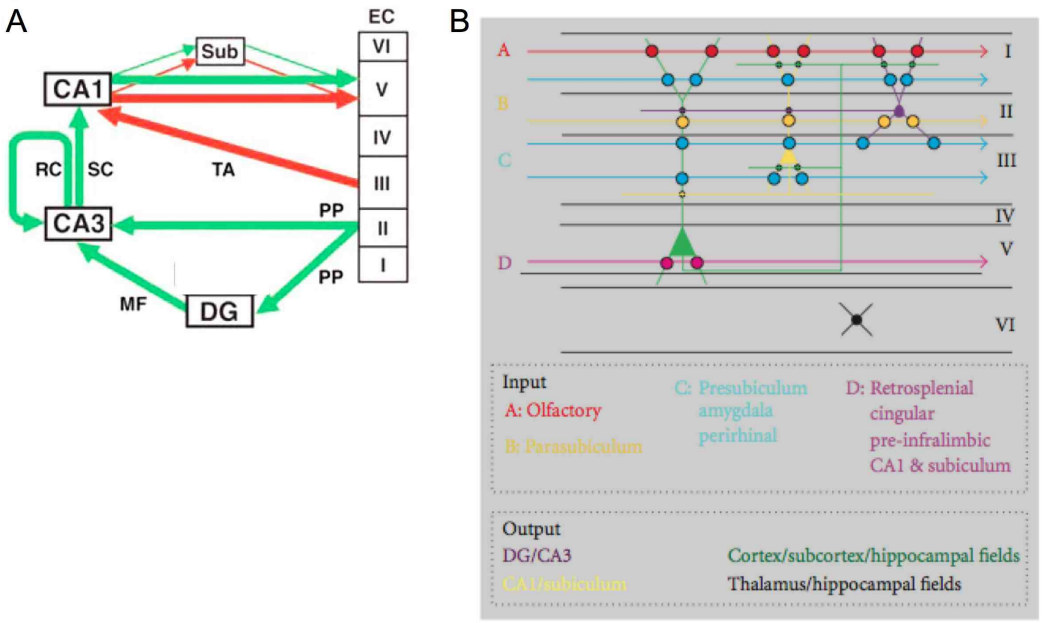


Figure 2.1: *MEC connectivity* (A) Circuit diagram of the parallel circuitry connecting the MEC and hippocampal subregions, the dentate gyrus (DG), CA3, CA1 and subiculum, including the trisynaptic pathway. Adapted from Nakashiba et al. (2008). SC: Schaffer Collaterals, MF: mossy fibres, RC: recurrent collaterals. (B) Schematic showing the different inputs and outputs for different MEC layers. Arrows indicate inputs. Circles indicate putative synaptic connections. Adapted from Canto et al. (2008)

2.3.2.2 *MEC layer II: the major input layer to the DG via the perforant path*

The most well studied layer of the MEC is the superficial layer II. It contains the highest proportion of grid cells (Sargolini et al., 2006) and contains the cells that show topographic organisation of properties along the dorso-ventral axis (Garden et al., 2008; Giocomo and Hasselmo, 2008). It also shows the strongest theta modulation of the layers (Burgalossi et al., 2011).

Cytoarchitecture Layer II can be easily distinguished using Nissl by the large darkly stained cells that run the full extent of the dorso-ventral axis. Cells in layer II are organised into hundreds of small patches (Burgalossi et al., 2011) that are highly metabolically activate and stain highly positive for cytochrome oxidase (Solodkin and Van Hoesen, 1996). They are often surrounded by myelinated fibres (Burgalossi et al., 2011). At the dorsal extent this layer thickens and extends into layer I (Amaral and Witter, 1989; Insausti et al., 1997) and the cells become smaller and paler, forming large patches (Burgalossi et al., 2011). Large axon bundles run along the dorsal surface of these patches (Burgalossi et al., 2011) , which may be the cause of changes in the reflective properties of the native tissue that creates a visualisable border.

Inputs to layer II cells The intriguing grid-like behaviour of layer II cells may be a function of the inputs they receive and compute. Few axon terminals impinge on layer II itself, except for parasubicular inputs (Caballero-Bleda and Witter, 1993), but the dendritic trees of layer II

cells often extend into layer I where many axons from neocortex and the deeper MEC layers terminate (Dolorfo and Amaral, 1998a; Burwell, 2000). Layer II cells receive inputs from the neurons in the large patches dorsal to the MEC, which have been termed centripetal axons (Burgalossi et al., 2011). Exploration of a novel environment causes expression of glycogen phosphorylase, a marker of activity that is probably glial cell based, in a modular manner in layer I (Walling et al., 2006). This may be of relevance to the inputs impinging on layer II cells, although the dendrites of cells in other layers also extend into layer I.

Projections from layer II: At least three cell types form the PP projection from layer II of MEC to the DG, but these are predominantly stellate, with fewer pyramidal neurons and interneurons (Schwartz and Coleman, 1981). Their axons synapse onto the spines of granule cells, forming excitatory asymmetric synapses, and onto inhibitory interneurons, forming excitatory and inhibitory connections (Witter, as cited in Canto et al. (2008)). These cells also send axon collaterals to CA3 and CA2 (stratum lacunosum- moleculare) and to the subiculum (Witter, as cited in Canto et al. (2008)), but it is not known if they form functional synapses there. The PP travels both ipsilaterally and contralaterally and connections to the contralateral MEC are made (Dolorfo and Amaral, 1998a). There is evidence that these contralateral connections arise from calbindin-positive principal neurons that have not been definitively characterised (Varga et al., 2010).

Cell types in layer II: Cell types in MEC are predominantly distinguished based on their morphology, although some molecular distinctions can also be made. The most prominent cell type of layer II is the stellate cell (Klink and Alonso, 1997; Gatome et al., 2010; Couey et al., 2013), which makes up approximately 74% of cells in the mouse and 66% in the rat (Gatome et al., 2010). Pyramidal cells are the second most prevalent cell type, accounting for 14% or 20% of cells in mice and rats, respectively (Gatome et al., 2010). Also found in layer II are cells with tripolar, bipolar, round and other appearance (Gatome et al., 2010), some of which likely map onto the non-spiny PP projection neurons described by Schwartz and Coleman (1981); Canto et al. (2008). Different cell types also vary in their arrangement, electrophysiological properties, the extent of their dendritic and axonal trees and their projections, which may provide clues about their different functions.

Stellate cells have received the most attention of the cells in layer II because of their prevalence, and the fact that their positioning and electrophysiological properties make them likely candidates for grid cells (see Pastoll et al., 2012). Stellate cells are typically located in the superficial or mid layer II. They are described as stellate because they typically have 6-10 primary dendrites extending from the soma (Klink and Alonso, 1997; Garden et al., 2008), rather than a single prominent primary dendrite, like pyramidal cells. They have spiny dendritic trees that extend throughout layers I-III and can occupy half of the medio-lateral extent of the MEC (Klink and Alonso, 1997), which is consistent with a lack of reported differences in properties throughout the medio-lateral extent. Their axons are relatively thick and project towards the angular bundle, sometimes from the base of the soma, and sometimes from a primary dendrite (Lingenhöhl and Finch, 1991). The axons occasionally branch throughout layers III-VI (Klink and Alonso, 1997). Paired recordings have demonstrated that stellate cells

show very low probability of synapsing on one another (Dhillon and Jones, 2000; Couey et al., 2013; Pastoll et al., 2013), which is surprising given the high overlap between their dendritic trees and local axon collaterals. However, glutamate uncaging in the superficial layers has revealed many functional synapses onto stellate cells from cells with superficial dendrites (Beed et al., 2010; Kumar et al., 2007), which suggests that stellate cells receive inputs either from other superficial cells or deep layer cells with superficially extending dendrites and axon collaterals.

The electrophysiological properties of stellate cells have been extensively reviewed by Pastoll et al. (2012), but several features are particularly relevant to the identification of cell type-specific molecules. Stellate cells can be identified by a prominent hyperpolarisation-induced sag, which is driven by the I_h current, which in turn is mediated largely by the Hcn1 channel (Nolan et al., 2007). Stellate cells have also been shown to be predominantly excitatory and therefore likely synthesise the transmitter glutamate, and it has been reported that some stellate cells express calbindin (Wouterlood, as cited in (Canto et al., 2008)). One particularly detailed anatomical study showed that cells that have a sag-like electrophysiological profile and stellate morphology express the protein reelin, while cells with a less pronounced sag profile, on average, express calbindin (Varga et al., 2010). Since stellate cells are so numerous and approximately 44-53% of all cells express each protein, but not the other (Varga et al., 2010), it seems likely that there may be two populations of stellate cells; one expressing reelin and one calbindin. This has important implications for mapping molecular expression to function, as reelin-positive cells have been shown to form the PP input to the DG, while calbindin-positive cells project to other regions, including the contralateral MEC (Varga et al., 2010). Only calbindin-positive cells are targeted by the cannabinoid receptor-expressing cholecystokinin-(CCk) positive interneurons present in layer II of MEC, and only these cells exhibit depolarisation-suppressed inhibition (Varga et al., 2010).

The second most prevalent cell type, the pyramidal cell, is typically positioned in the deeper extent of layer II (Alonso and Klink, 1993), but the prominent primary dendrite of pyramidal cells and its branches also extend throughout layers I-III. There is evidence from glutamate uncaging studies that the neurons synapsing onto pyramidal cells are more proximal to the cell body, and more likely to be deep layer neurons, than those synapsing onto stellate cells. However, since glutamate uncaging can be ambiguous when dendrites are very distant from the cell body, this requires confirmation using more direct approaches (Beed et al., 2010).

The other types of cell in MEC layer II include non-spiny projection neurons (Canto et al., 2008) and inhibitory neurons. Basket cells, which express the calcium-binding protein parvalbumin, which is expressed at high levels in the neuropil of MEC, may be particularly important for enabling communication between stellate cells (Jones and Buhl, 1993; Pastoll et al., 2013). These cell types are summarised in Table 2.1. Essentially, although research has identified different morphologies, markers, projections targets and electrophysiology, it is not yet clear how the following properties interact to form a coherent picture of the different cell types in layer II: (1) the cells providing the input to the dentate gyrus, (2) cellular morphology, (3) grid cell behaviour, (4) cell physiology, and (5) molecular expression.

Cell type	Defined by	Inputs	Projection	Molecular expression	Electro-physiological profile	Morpho-logical profile
Stellate	morphology	LII Pv interneurons	DG*	Calb*[3], Hcn1+	Fast, prominent, hyperpolarisation-induced sag[4]	stellate
Pyramidal	morphology		DG*		Lack of fast, prominent hyp-induced sag[4]	pyramidal
Reelin+[6]	molecule		DG	Reelin	Sag*	Stellate^
Calbindin+[6]	molecule	LII Cck interneurons	Contra MEC	Calbindin	Sag*	
Head-direction	Function		Hippocampus ^[7]			
Grid	Function		Hippocampus ^[7]			stellate*/ pyramidal*[1]
Conjunctive	Function					
Border	Function		Hippocampus ^[7]			
	Projection	Dorsal parasubicular patches[2]				

Table 2.1: *Summary of cell types* Key: ∗ some, ◦ weak evidence, + some more than others. (1) (Domnisoru et al., 2013), (2) (Burgalossi et al., 2011),(3) (Canto et al., 2008),(4) (Klink and Alonso, 1997),(6) (Varga et al., 2010), (7) (Zhang et al., 2013).

2.3.2.3 *MEC layer III: the major input layer to CA1*

Cytoarchitecture: Layer III of the MEC contains cells that project to the CA1 region of the hippocampus, where they form connections critical for place field stability and spatial memory consolidation (Remondes and Schuman, 2004; Brun et al., 2008a). Layer III can be distinguished from layer II using Nissl stain by the lack of darkly stained large neurons and presence of a sparser, wider column of large-medium cells that are stained less intensively. Layer III is bordered on the anterior and medial side by the laminar dissecans, or layer IV, a layer predominantly containing fibres with few cell bodies.

Cell types in layer III Layer III cells are predominantly pyramidal, although many other cell types exist. Pyramidal cell dendrites show similar branching throughout layers II and I to layer II cells, and may therefore receive similar inputs. However, since most inputs to MEC terminate in layer III, they also receive the major presubicular input, which targets both pyramidal cells and interneurons (van Haeften et al., 1997), directly, and exerts an overall inhibitory effect (Tolner et al., 2005). Pyramidal cells are strongly innervated by local inhibitory interneurons (Wouterlood, as cited in Canto et al. (2008)). As well as being the only MEC cells to target proximal CA1, pyramidal neurons also target the distal subiculum, CA2 and CA3. In fact, CA3 inputs and MEC inputs target the same CA1 pyramidal cells and interneurons with high incidence (Kajiwara et al., 2008). Layer III cells also project to infralimbic and olfactory cortex (Insausti et al., 1997), and the large patches dorsal to the MEC (Burgalossi et al., 2011).

The organisation of inputs to MEC layers II and III shows a high degree of similarity due to the similarity of dendritic coverage. To understand what distinguishes the computations of these layers, thus providing different information to different hippocampal structures, it will be necessary to identify just how different the synaptic connectivity is between layer II and III cells, and map this to differences in molecular expression and physiology.

2.3.2.4 *MEC layers V & VI: the major projection layers*

Unlike layers II and III, MEC layer V and VI predominantly receive hippocampal input rather than providing input to the hippocampus (Swanson and Cowan, 1977; Tamamaki and Nojyo, 1995), which suggests that they may exhibit different characteristics to the superficial layers. However, layer V pyramidal neurons have dendrites that extend through the superficial layers, thus enabling them to receive similar inputs to neurons in superficial layers. The extent to which they do has not been fully elucidated. Layer V neurons show characteristics that have implicated them in working memory function (Egorov et al., 2002). Although grid cells are found in the deep layers, they are less common and their firing properties differ (Sargolini et al., 2006). There is currently uncertainty regarding how active layer V neurons are during novel exploration, with somewhat contradictory data published by different groups (Burgalossi et al., 2011; Sargolini et al., 2006). In later chapters I will show that the expression of an immediate early gene following novel exploration changes more in layer V/VI cells than in other layers, so I will outline the properties of these layers in slightly more detail, and discuss possible reasons for the differences alluded to.

Cytoarchitecture: Layer V exhibits a stratified structure that can be subdivided into a superficial layer bordering the laminar dissecans, and deeper layers. The superficial layer contains large-medium pyramidal cells that, like layer II cells, stain darkly with Nissl (Canto et al., 2008); this has been and will be referred to as layer Va. The deeper layers (sometimes referred to as layers Vb and Vc) contain smaller pyramidal cells that are densely packed; this will be referred to as Vb. Layer VI contains smaller, more densely packed cells and is delineated by the white matter but cells intermingle with layer V neurons, making it difficult to distinguish. The lack of precision distinguishing these regions highlights the difficulty in the MEC of associating particular cell types with function. In sagittal sections, at the dorsal extent, the layer V/VI border is particularly difficult to establish, thus emphasising the need for molecular markers of cells.

Inputs to layers V and VI: The deep layers of MEC receive topographical inputs from the proximal (predominantly) CA1 region and distal subiculum and although other MEC layers also receive a projection from CA1 (Melzer et al., 2012), this is considered the dominant projection (Tamamaki and Nojyo, 1995). Projections from the subiculum are ipsilateral (Kloosterman et al., 2003), are capable of plasticity (Craig and Commins, 2006) and predominantly target layer Va (Kloosterman et al., 2003). CA1 inputs are also ipsilateral (Swanson and Cowan, 1977) and capable of plasticity (Craig and Commins, 2007). Aside from these inputs, the deep layers receive direct inputs from cingulate cortex, particularly the retrosplenial area and infralimbic and prelimbic areas (Jones and Witter, 2007; Sugar et al., 2011). Lesion studies

suggest the RSC is necessary for path integration-based navigation and topographic memory (Sutherland et al., 1988), indicating that layer V may be an important site for the integration of processed hippocampal inputs and contextual information.

Projections from layer V: Most cortical projections from MEC arise in the deep layers and terminate in visual and retrosplenial cortex; other EC-cortical projections mainly arise from lateral EC (Insausti et al., 1997). Layer V cells also project to the amygdala, although this appears to be a very limited range of ventral layer Va neurons (McDonald and Mascagni, 1997). Although the major hippocampal input is from the superficial layers, layer IV-VI projections also terminate in the inner molecular layer of DG (Deller et al., 1996). Although the projection is small, activation at the LEC/MEC border is sufficient to activate the DG (Koganezawa et al., 2008), but the significance of MEC deep layer cells to this remains unclear due to the low spatial resolution of the imaging used.

Cell types in layers V and VI: There are several different principal cell types in layer V, which included pyramidal cells, horizontal cells and polymorphic cells (Hamam et al., 2000). Although morphologically distinct from one another, these cell types are electrophysiologically heterogenous and show almost as much variability within a morphological classification as between (Hamam et al., 2000), which suggests that elucidating functional differences may require a more high dimensional analysis of features.

Pyramidal cells in layers V and Va possess a large, long apical dendrite that penetrates the superficial layers of MEC and branch extensively in layer I (Hamam et al., 2000). The apical dendrites of layer VI pyramidal neurons also follow this course, but not always (unpublished data, cited in Canto et al., 2008). These are the only deep layer cells that receive inputs that terminate in the superficial layers. The basal dendrites spread widely throughout the other deep layers, whereas axons often travel towards the angular bundle or superficial layers (Canto, as cited in Canto et al., 2008). Layer V pyramidal dendrites are spiny and they receive both distal and proximal inputs (Hamam et al., 2000; Gasparini, 2011) (Medinilla et al., 2013), some of which are known to originate in the presubiculum and parasubiculum (Canto et al., 2012). The size of the cell body of pyramidal cells varies and there is some evidence that smaller pyramidal cells have larger and more densely distributed spines than large cells (Hamam et al., 2000). Horizontal cells do not have a long apical dendrite extending into the superficial layers, but instead their basal dendrites extend horizontally. Spines are sparsely distributed along the dendrites and are most abundant on the apical dendrite (Hamam et al., 2000). Horizontal cell axons always travel towards the angular bundle having branched extensively within the deep layers. A number of other types of multipolar, fusiform and bipolar cells are also found throughout the deep layers (summarised in Canto et al., 2008)

Most of the attention that has been directed towards layer V cells relates to how their electrophysiological profile might provide a neural basis for working memory in the system.

2.3.2.5 *Summary of cell types and connections:*

Much of the focus of the entorhinal-hippocampal system has been dominated by the view that layer II cells project to the dentate gyrus, while deep layer cells receive inputs from the

hippocampus (Witter). To achieve a holistic view of the function of the system will likely require an understanding of why the deep MEC layers receive inputs from the superficial layers and from other cortical regions that project to them, why there is so much intraconnectivity within the superficial layers, yet the most dominant cell type, the stellate cells, do not connect to one another at all (Dhillon and Jones, 2000; Couey et al., 2013; Pastoll et al., 2013), and what the significance of the parallel pathways from layer II into the DG and layer III into CA1 (Dolorfo and Amaral, 1998b), with additional inputs to CA2 and CA3 is.

2.3.3 *The dorso-ventral gradient in properties of the MEC*

One of the most intriguing features of the MEC, and one that increases its interest as a system for probing its molecular profile is that properties of cells that are considered to be of the same type vary along the dorso-ventral axis of MEC. These properties extend from the high-level representation of space encoded by the grid cells (Hafting et al., 2005) to the density of specific ion channel currents (Garden et al., 2008). A critical question for the field is what maintains these differences in properties between dorsal and ventral cells.

2.3.3.1 *Grid cell size and spacing*

When grid cells were first discovered, but were yet to be described as such, it was thought that robustly modulated cells only existed in the most dorsal region of MEC (Fyhn et al., 2004). In fact, the earliest investigations into the presence of spatially modulated cells identified only weak spatial modulation, because cells were being selected from intermediate regions (Barnes et al., as cited in Fyhn et al. (2004)). It was soon realised that grid cells do exist throughout the dorso-ventral extent but that observations of the organisation of firing fields in ventral cells depend on the use of larger (2m diameter) environments (Hafting et al., 2005). In both rats and mice (Fyhn et al., 2008), the spacing and size of grid fields in dorsal cells is smaller than those of ventral cells. The low number of cells recorded from in each animal, followed by pooling of data across animals, led to initial suggestions that the increase in size from dorsal to ventral cells occurs gradually (Hafting et al., 2005; Brun et al., 2008b). However, a vast project to collect data from many cells in each animal, and cluster it based on firing properties, revealed that the changes occur along a modular, discretised gradient (Stensola et al., 2012). Cells along extensive areas of the dorso-ventral extent can show similar-sized fields within the same animal, but at different dorso-ventral locations the probability that a cell has a grid field of a particular size changes. Dorsal cells are all likely to have similar-sized fields, whereas ventral cells may have fields of different sizes (Stensola et al., 2012). The finding of this gradient suggests that the MEC is differently organised along this axis, but the extent to which intrinsic properties of cells and differences in connectivity contribute to this topography remains unclear.

2.3.3.2 *Layer II cell morphology and organisation*

In the years since the discovery of a gradient in the spacing of grid fields, researchers from several labs have identified changes in the properties of cells along this axis in MEC that

could underlie the differences in grid cell size and spacing. One of the varying properties is stellate cell morphology along the axis, including a reduction in the total dendritic surface area due to a decrease in the number of primary dendrites (Garden et al., 2008). The size of the stellate cell body is also smaller for more ventral cells (Garden et al., 2008; Burgalossi et al., 2011). One study that analysed Nissl-stained sections to estimate cell type, number and size showed that this reduction was predominantly true for a subgroup of stellate cells, the polygonal stellate cells, and to a lesser degree, was also true for pyramidal cells in both mice and rats (Gatome et al., 2010). Cells in layer II are organised into patches that can be seen in both Nissl and using cytochrome oxidase staining (Burgalossi et al., 2011). Along the dorso-ventral gradient the size of the patches increases and there is also evidence that the degree of myelination decreases (Burgalossi et al., 2011). Although these results are intriguing, it is difficult to predict how they could map onto changes in grid behaviour. Two possibilities are that dorsal cells receive a larger number of inputs, and that these inputs may arise from cells in a more confined region, thus providing higher spatial resolution.

2.3.3.3 *Changes in the tuning of synaptic integration by layer II stellate cells*

Morphological differences in the dorso-ventral organisation of properties are accompanied by differences in the way that stellate cells integrate incoming inputs (Garden et al., 2008). This feature could be strongly influenced by the expression of certain molecules. Spontaneous, non-NMDA-mediated EPSPs [4] in stellate cells show a slower rise time and half-width with increasing distance from the dorsal border (Garden et al., 2008). As a result, dorsal neurons have a narrower time window for the detection of coincident inputs, and are less likely to sum high frequency inputs (Garden et al., 2008), such as the gamma frequency synaptic activity observed in behaving animals in vivo (Chrobak and Buzsaki, 1998). These results are not explained by changes in synaptic transmission in dorsal and ventral cells, but are accompanied by a steep increase in cell input resistance, membrane time constant and current threshold and a decrease in the amplitude of the HCN1-dependent sag away from the dorsal border (Garden et al., 2008). After simulations demonstrated that morphological differences were unlikely to account for the steep increase in input resistance, but instead they were likely due to a reduction in membrane conductance, the authors found that this reduction was likely due to a decrease in the current density of the HCN1-mediated Ih current and the quinidine-sensitive leak K+ current, which are hyperpolarising or depolarising, respectively (Garden et al., 2008). These results provide clear predictions for molecular differences that could underlie the electrophysiological behaviours, and suggest that a more systematic molecular analysis could reveal further contributors.

2.3.3.4 *The relationship between the dorso-ventral gradient in MEC and the septo-temporal gradient in the hippocampus.*

The potential functional significance of dorso-ventral changes in grid cell properties is increased by the evidence of similar topographical changes in hippocampal cell function along its corresponding septo-temporal axis. Place field size increases with temporal extent (Jung

4 Spontaneous EPSPs occurred in the presence of D-APV to block NMDA receptors and picrotoxin to block GABAA receptors, and were abolished by 5 mM NBQX, indicating that they were mediated by AMPA/kainate receptors.

et al., 1994), while the probability that lesions disrupt spatial learning and recall decreases (Steffenach et al., 2005; Kjelstrup et al., 2002). In fact, lesions to ventral hippocampus have no effect on performance on spatial tasks such as the water or radial maze (Jarrard et al., 2012), which has led to suggestions that dorsal and ventral hippocampus are functionally distinct (Fanselow and Dong, 2010; Hock and Bunsey, 1998; Royer et al., 2010).

The evidence the MEC and hippocampus show similar organisation of properties is supported by the anatomical relationship between them. Ventromedial MEC cells are more likely to target temporal DG whereas dorsal MEC (CE) cells target septal DG, both in rats and in mice (Dolorfo and Amaral, 1998b; van Groen et al., 2003) (see Fig. 2.2). The converse relationship is also true in that septal CA1 cells target dorsal MEC whereas ventral CA1 cells target ventral MEC (Cenquizca and Swanson, 2007). In addition, topography is maintained throughout the circuit in that recurrent loops in hippocampus are spatially restricted (van Strien et al., 2009).

Together, these changes in anatomy and physiology may play an important role in the encoding of spatial environments. To elucidate this, it may be necessary to selectively modulate these features, which will require fine molecular control.

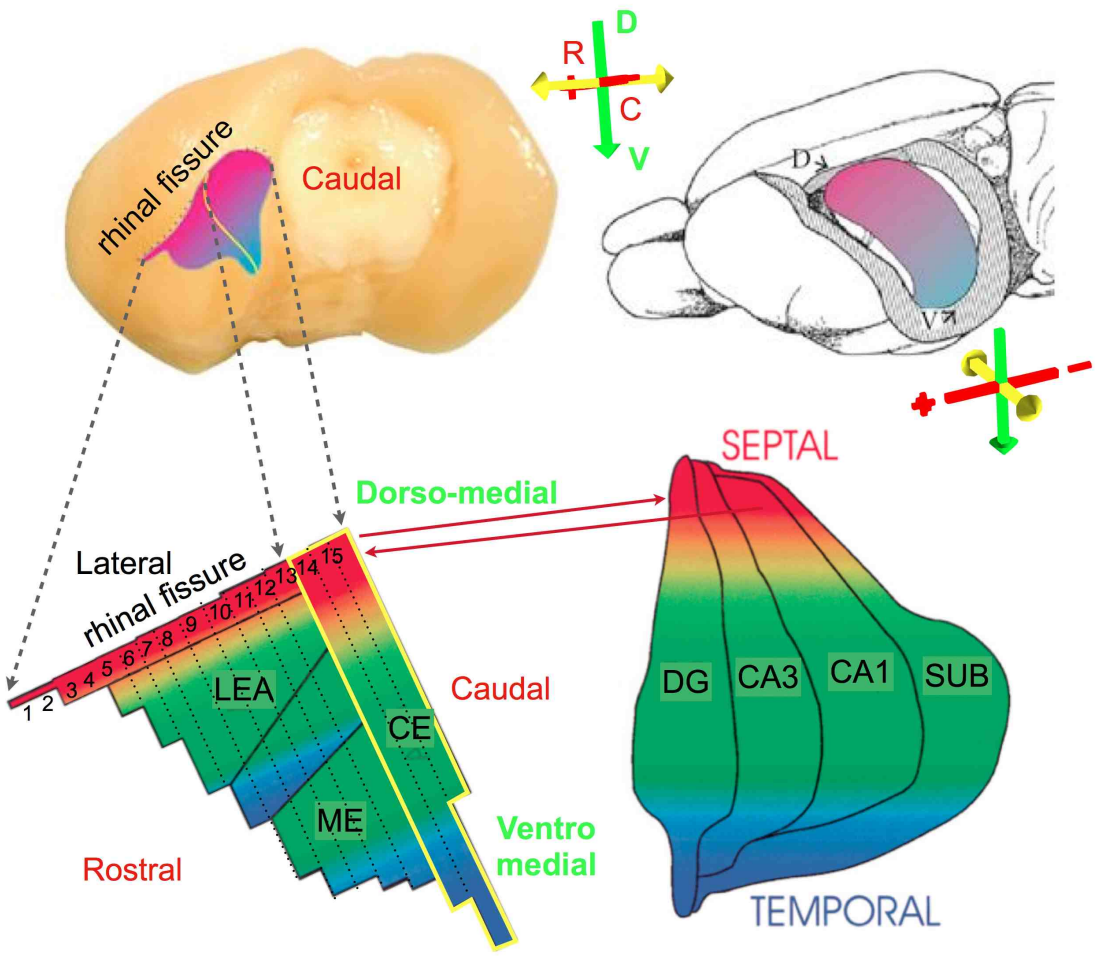


Figure 2.2: *Topography in the entorhinal-hippocampal circuit* Images adapted from Canto et al. (2008) indicate the positioning of the MEC (caudal) and LEC (lateral). The dorsolateral/dorsomedial region is marked in pink/red whereas the ventromedial region is marked in blue. Arrows indicate the orientation of the brain in 3D. Below are figures adapted from van Groen et al. (2003), which reveal the topographic organisation of the projection from areas CE (caudal MEC), ME (ventromedial MEC) and LEA (the lateral EC) to regions of the hippocampus. Red regions indicate dorsolateral/dorsomedial/septal regions and Blue regions indicate temporal regions. On the left is an unfolded map of the entorhinal area, developed using coronal sections (numbered from rostral to caudal). The 'dorsolateral' region of MEC is in fact positioned at close to the most medial extent.

2.3.4 *The vulnerability of MEC to disease*

2.3.4.1 *Alzheimer's Disease pathology is first evident in MEC*

In Alzheimer's Disease (AD) the MEC is the one of the first locations in which degeneration is first evident (Braak and Braak, 1991). AD is characterised by the presence of extracellular amyloid deposits and intraneuronal neurofibrillary changes that occur as a result of tau accumulation. The latter provide the best neuropathological correlate of clinical symptoms

(Braak and Braak, 1991; Duyckaerts et al., 2009) and first develop in the early stages of the disease in transentorhinal cortex and layer II of human medial entorhinal cortex (Braak and Braak, 1991). There is a profound loss (~ 60%) in the number of layer II neurons even in mild AD, and up to 90% of layer II cells are lost in severe AD (Gómez-Isla et al., 1996). Atrophy and abnormal metabolism in patients with mild cognitive impairment is also evident using structural and functional MRI (reviewed in Masdeu et al. (2005)). Although the reason for the vulnerability of the EC to AD is unclear (see (Harris et al., 2010)), there is evidence that disruption of Aβ signalling [5] through overexpression of human APP in mice may transynaptically disrupt entorhinal/hippocampal circuitry (Harris et al., 2010). Mice with transgene overexpression of APP in superficial EC and parasubiculum exhibit disinhibition, hyperactivity and age-related cognitive deficits in memory, which correspond with levels of abnormal Aβ deposits at PP terminals in older mice and alterations in synaptic-activity-related molecules in the DG granule cells (Harris et al., 2010). A similar distribution of abnormal cellular activity is observed in mice expressing pathological human tau (Ramirez et al., 2011), which forms the neurofibrillary tangles. These mice show working memory deficits (Ramirez et al., 2011).

Several key questions arise from these results: (1) what makes the EC so vulnerable to AD pathology; is it a function of circuit organisation, or do combinations of genes expressed in these cells interact to make them vulnerable?, (2) what can the spread of pathology confer about the entorhinal/hippocampal circuitry, and (3) can the selectivity of AD pathology and functional deficits aid our understanding of the functions of microcircuits within MEC? Of interest, research has shown that although not central to the disease, other neurodegenerative diseases, such as Parkinson's disease, Pick's disease and Huntington's disease, also show pathology in MEC (Braak et al., 2000). It is therefore possible that these diseases share common disrupted pathways that are particularly vulnerable in the MEC.

2.3.4.2 *Oscillatory and cellular disruption in schizophrenia*

AD is not the only disease to be associated with abnormalities in MEC; in schizophrenia, there is some evidence that the neuron clusters of layer II are poorly formed and the layers are disorganised, superficial neurons are smaller and the microtubule-associated protein MAP2 is abnormally expressed (summarised by Arnold (2000)). There have also been reports of aberrant glutamatergic and catecholaminergic innervation (summarised by Arnold (2000).) ISH and single-cell transcriptome analysis (Hemby et al., 2002) have revealed abnormal mRNA expression of various transcription factors, ion channels, and neurosecretory pathway-related proteins in stellate neurons (summarised by Arnold (2000)). The genes affected include growth-associated protein-43 and complexin II (Hemby et al., 2002). In addition, the synaptic genes Nr1, GluR3, synaptophysin, Snap23 and Snap25 are all downregulated in the stellate cells of schizophrenia patients compared with controls. However, there is no evidence of neurodegeneration (Arnold et al., 1991), suggesting that the vulnerability of MEC to the two insults may be caused by different factors. Perineuronal net formation, which regulates cell function, is also disrupted in schizophrenia (Pantazopoulos et al., 2010).

5 this is likely to be a major factor in AD pathogenesis (Glenner & Wong, as cited in a review by (Tanzi and Bertram, 2005))

We remain relatively in the dark about how and why these abnormalities develop and how they impact on EC function. Patients with schizophrenia perform poorly on WM tasks, which could potentially be due to MEC damage (de la Serna et al., 2009). In an animal model of schizophrenia, the LPA1-deficient mouse, there is a decrease in gamma oscillations in MEC and a corresponding reduction in total GABA- and parvalbumin-containing neurons in layer II, which has led to suggestions that abnormal function of interneurons may underlie network function (Cunningham et al., 2006), but this will require validation across other animal models.

2.3.4.3 *Altered excitability and connectivity in MEC in models of epilepsy*

Temporal lobe epilepsy (TLE) is the most common form of epilepsy found in humans (reviewed in Engel (1996)) and has also been strongly associated with pathology in MEC. Layer III neurons have been particularly strongly implicated because tissue extracted from patients shows pronounced cell loss in layer III, and to a lesser degree in layer II (Du et al., 1993). In an animal model of TLE in which seizures are initially induced using kainic acid, which causes the animal to develop a chronic form of epilepsy, layer III pyramidal cells remain spontaneously active over 24 hours after the induction of TLE, despite no overt seizures (Shah et al., 2004). This is accompanied by a reduction in I_h and in immunoreactivity for the Hcn1 and Hcn2 channels that mediate it (Shah et al., 2004). TLE is also associated with the PP projection in animal models, and in particular to hyperexcitability of stellate cells (Tolner et al., 2005). Unlike in layer III, this appears to be due to loss of GABAergic input and reduced inhibition, but it does not appear to result from direct hyperexcitability from layer III cells (Kumar et al., 2007).

The pathology evident in MEC that is associated with schizophrenia and epilepsy suggests that inhibition within the superficial layers plays a critical role in maintaining the stability of the system. Elucidating how information flows within and through the superficial layers will provide insight into why disrupting the excitatory/inhibitory balance is so detrimental.

2.3.5 *Summary*

The entorhinal-hippocampal circuit is an intriguing system that, somewhat ironically, can seem difficult to navigate. There is an overwhelming amount of data regarding potential connections, potential functions and intrinsic properties. At some point this will be consolidated into an explanatory, mechanistic model of the system that encompasses all its major features, but for the time-being, this seems rather far away. As discussed in the first section, genetic descriptions and connectivity descriptions can be used in a complementary way to advance knowledge in a field, but in the entorhinal system, with a few exceptions, this approach has not been fully exploited.

In the next section, I will discuss the characteristics of immediate early genes, which are often used as markers of activity, and consider how this information could be used to broaden our understanding of the entorhinal-hippocampal system and its function in representing space-related experiences.

2.4 ACTIVITY-DEPENDENT GENE EXPRESSION

In the preceding sections I have outlined the importance of gene expression for distinguishing different types of cells and their function, and examined how an understanding of gene expression could further inform our understanding of the anatomy and function of the MEC. Many years ago it was discovered that the ability of neural activity to cause long-term modifications of synapses is blocked by protein synthesis inhibitors (Flexner, 1963, as cited in (Lyford et al., 1995)). This led to research demonstrating that this protein synthesis, which likely underlies the extensive capacity and specificity of information storage in the brain, depends on the dynamic and neural-activity related regulation of the expression levels of certain genes. Activity-dependent gene expression may therefore provide critical insight into the circuit function and learning capacity of the MEC.

2.4.1 *An introduction to activity-regulated immediate early genes*

Immediate early genes (IEGs) are genes that are actively transcribed without the requirement for protein synthesis. The most well-studied genes, the activity-regulated cytoskeleton-associated protein, *Arc,* Fos and Egr1 (zif268), are commonly used as indicators of neuronal activity (Flavell and Greenberg, 2008). IEGs fall into two categories: DNA-binding proteins, such as transcription factors, and effector molecules. While Fos and Egr1 are transcription factors that go on to drive the expression of late-response genes, *Arc* is a cytoskeletal protein that directly influences synaptic function (Steward and Worley, 2001b). The IEGs show a variety of features and vary in the type and specificity of stimulus that activates them, the upstream pathway mediating their activation and their functions within the cell. They initiate a cascade that can alter the expression levels of hundred of other genes, for example following memory tasks (de Quervain and Papassotiropoulos, 2006; Rapanelli et al., 2010; Klur et al., 2009; Miyashita et al., 2009).

Factors that up-regulate immediate early genes The principle factors that influence IEGs *in vitro* are neuronal activity and growth factors, while behavioural stimuli include novelty, stress, seizures and the requirement for learning (Bramham et al., 2010). For example, LTP protocols such as high-frequency stimulation, or electrically induced seizures stimulate a large number of IEGs, often at different times scales ((Nedivi et al., 1993; Yamagata et al., 1993), reviewed by (Leslie and Nedivi, 2011)), whereas specific types of learning can induce a more select number of IEGs that are specific to that particular stages learning (von Hertzen and Giese, 2005). Although IEGs such as *Arc* can be induced by LTD, stimuli that typically cause LTD, such as low frequency stimulation, do not increase *Arc* transcription or synthesis (Lyford et al., 1995). IEGs, in this case *Arc*, can also be induced *in vivo* by 5HT receptor activation (de Foubert et al., 2007).

IEG function Activity-dependent genes have been linked to long-term physiological and structural changes (reviewed in (Leslie and Nedivi, 2011)) that are driven by changes in level of gene expression. These include structural and compositional changes that lead to the

strengthening, weakening, addition and elimination of synapses, changes that are required in various forms of plasticity, including NMDA-dependent LTP (Leslie and Nedivi, 2011). The IEG that has been most clearly linked to changes during NMDA-dependent LTP is *Arc* (Lyford et al., 1995), also known as *Arg3.1* (Link et al., 1995). Since the discovery that plasticity is input-specific, many researchers sought to identify the mechanism that enables proteins to be selectively trafficked to active synapses. It emerged that *Arc* mRNA, once induced and transcribed in the cell nucleus, is then trafficked to dendrites and stored at sites of synaptic activity where it can be locally translated (Lyford et al., 1995; Rodríguez et al., 2005). This differs from *Egr1/zif268*, which accumulates around the cell body (cited in Bozon et al., 2002). *Arc* is now known to be critical for a number of different types of plasticity and spatial learning (see (Bramham et al., 2010) for a review) and its control mechanisms have also been elucidated (Kawashima et al., 2009; Pintchovski et al., 2009). It is therefore an attractive candidate for elucidating circuit mechanisms within MEC.

2.4.2 *Arc as an indicator of cellular and circuit function*

Arc has received a huge amount of attention since its discovery in 1995 (Lyford et al., 1995; Link et al., 1995). The aim of this section is not to provide a comprehensive overview of *Arc* - several reviews have focussed on different properties of *Arc* (Tzingounis and Nicoll, 2006; Bramham et al., 2010, 2008) - but to concentrate on properties specific to this thesis. I will therefore outline the *in vivo* conditions under which *Arc* is stimulated, with an emphasis on the up-regulation of *Arc* following spatial memory-related tasks. I will then discuss factors that influence transcription and translation, and how closely these processes parallel one another. I'll consider the effect of disrupting Arc transcription or translation on cellular and behavioural processes. Finally I'll outline associations of *Arc* with disease states.

2.4.2.1 *Key features of the Arc protein*

Arc encodes a cytoskeletal protein that is found throughout the neuron. Its baseline dendritic expression is non-zero and can be abolished by applying the NMDA antagonist MK801 (Lyford et al., 1995), suggesting it functions in resting circuits. ISH data has revealed that RNA levels are particularly higher at baseline in regions of the brain including the reticular hypothalamic nucleus, the lateral olfactory tract nucleus, CA1 and limbic and orbital cortex (Ons et al., 2004).

Arc RNA and protein is found at particularly high levels in the dendrites and post-synaptic density where it associates with F-actin, possibly resulting from homologous regions with α-spectrin (Lyford et al., 1995). What makes *Arc* particularly unique as an IEG is that synaptic activity causes its mRNA to be actively transported from the nucleus to active synapses where is locally translated. *Arc* mRNA can be detected in distal dendrites within 1 hour of transcription (Lyford et al., 1995; Steward and Worley, 2001a). This transport process has received considerable attention. The ability of *Arc* to move out of the nucleus is conferred by a 3' UTR dendritic targeting region (Kobayashi et al., 2005), but other binding sites enable trafficking once in the cytoplasm (Steward and Worley, 2001a). There is evidence that *Arc* is packaged into RNA granules with αCaMKII, with which it shares an RNA sequence in the

coding region that enables dendritic targeting (Gao et al., 2008). *Arc* protein also accumulates in the nucleus but its function there is not known (Bramham et al., 2010). These features are summarised in Fig.2.3

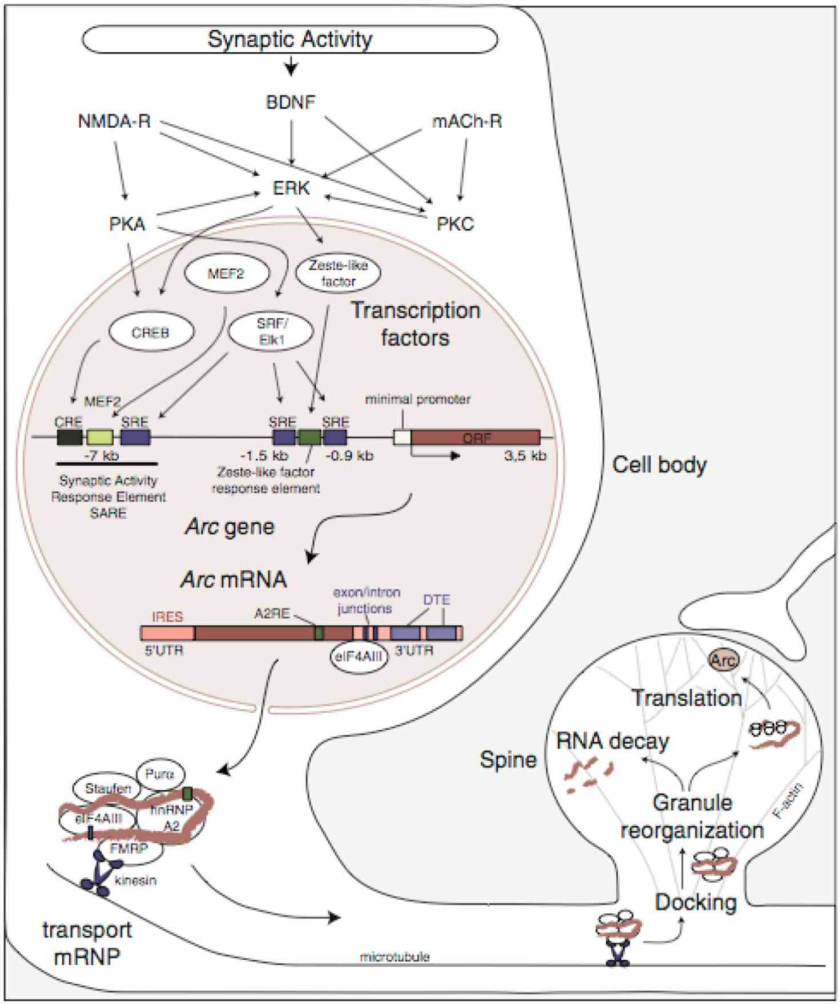


Figure 2.3: *Key features of Arc* Schematic showing the transcriptional regulation of Arc following synaptic activity, its transport, and localisation to synapses. Adapted from (Bramham et al., 2010) Fig.1.

2.4.2.2 *What types of behavioural in vivo circuit activity drive changes in Arc expression?*

Arc has been shown in various *in vitro* slice, culture and *in vivo* conditions to be up-regulated by many different stimuli - for a review see (Bramham et al., 2010). The natural stimuli that stimulate *Arc* transcription and protein expression appear to fall under the major categories of novelty (Guzowski et al., 1999), stress (Ons et al., 2004), environmental enrichment (Pinaud et al., 2001), and direct requirement for learning (Guzowski et al., 2001), but the relative importance of these features is not well understood, as they are difficult to dissociate. Moreover, it is not necessarily the case that transcription and translation parallel one another (cf Guzowski et al. (2006); Kelly and Deadwyler (2003)). ISH has been used to show that exposure to a novel environment, which may also increase stress and prime circuits for learning, acti-

vates Arc RNA in the hippocampus (Guzowski et al., 1999; Ons et al., 2004), and particularly strongly in limbic, cingulate and orbital cortex, claustrum, dorsal endopiriform nucleus and lateral septum of rats (Ons et al., 2004). Genetically modified knockin mice that express GFP under the control of the *Arc* promoter have shown that *Arc* expression is also increased in the visual cortex (Wang et al., 2006). Following novel exploration, *Arc* RNA in rat hippocampus is expressed at similar levels following exposure to novel and familiar environments if the familiar environment has not been experienced in the recent few hours (Guzowski et al., 2006), which suggests that complete novelty is not a requirement for expression. Performance on water maze tasks also stimulates *Arc* expression in CA1 pyramidal neurons whether there is a requirement for new learning or not (Gusev et al., 2005; Guzowski et al., 2001), but newly trained animals experience more up-regulation than over-trained animals, both on the water maze and on lever-pushing tasks (Kelly and Deadwyler, 2002; Guzowski et al., 2001). It's interesting to note that these experiments suggest opposing correlations between performance and *Arc* expression, with Kelly and Deadwyler (2002) demonstrating that low performance is associated with high *Arc* expression, whereas (Guzowski et al., 2001) reported higher *Arc* RNA expression in the best performers. These differences could reflect differences in motivation and in the specificity of the task to the function of the region studied, CA1.

Stress also appears to impact directly on *Arc,* in the absence of a novelty or learning component. For example, sleep deprivation increases *Arc* RNA expression in neocortex, hippocampus, piriform cortex and amygdala, which remains high even if the animals are sacrificed after 4 hours recovery sleep (Thompson et al., 2010).

Specificity of Arc up-regulation in hippocampal-entorhinal-related tasks: Many of the observations regarding *Arc* up-regulation are pertinent to the understanding of the requirements of the entorhinal-hippocampal system. Notably, the majority of experiments are performed on hippocampal neurons. Experiments have shown that tasks can selectively up-regulate Arc within subregions of this system (Hartzell et al., 2013; Gusev et al., 2005), that Arc shows more specificity than other immediate early genes for the requirement of a spatial task (Guzowski et al., 2001), that *Arc* expression and neural activity in CA1 are linked (Guzowski et al., 1999, 2006), and that *Arc* is predictive of the maturation of newly formed neurons in the dentate gyrus, where its functions may precede the formation of glutamatergic synapses (Kuipers et al., 2009).

Arc expression and activity in pyramidal cells in CA1 are closely but not directly correlated (Guzowski et al., 1999). When an animal experiences different environments there is evidence from tetrode recordings that place fields remap and that distinct but overlapping populations of CA1 place cells are recruited (Wilson and McNaughton, 1993). To test whether *Arc* transcription across a network of neurons follows a similar pattern, Guzowski et al. (1999) developed a highly sensitive catFISH (cellular compartment analysis of temporal activity by fluorescent ISH) technique. CatFISH involves high-resolution imaging of ISH to measure changes in *Arc* levels independently in the nucleus and cytoplasm. This enables the specificity of Arc up-regulation to a recent behavioural experience to be assessed. Intranuclear foci of *Arc* RNA appear in the nucleus after 2 minutes but return to baseline by 16 minutes as the RNA moves to the cytoplasm. Guzowski et al. (1999) showed that different

populations of neurons that are active in different environments selectively express *Arc* in the part of the neuron that corresponds with the timing prior to sacrifice that the rat was exposed to each environment, indicating that transcription is a good indicator of neural activity. Perhaps most interesting, the relationship between neural activity and *Arc* transcription is not constant under conditions of repeated exposure to the same environment; while neural activity is maintained over repetitions, Arc transcription is suppressed, indicating that the relationship is plastic and modified by behavioural experience (Guzowski et al., 2006). Further evidence for a dissociation between activity and transcription is that *Arc* is induced by 1-trial learning in CA3 and CA1 neurons of the hippocampus, but only in the presence of an intact theta input from the medial septum (Miyashita et al., 2009), which is critical for hippocampal LTP and learning (McNaughton et al., 2006b), but not location-specific firing (Mizumori et al., 1989). Thus, *Arc* does not necessarily indicate place field activity, but the process of learning.

The sensitivity of *Arc* expression to changes in environment is not uniform across the hippocampus. Following sequential exposure to two different environments, *Arc* is more likely to be up-regulated in different pyramidal cells in the proximal region of CA1 than in the distal region (Hartzell et al., 2013). Since the proximal region receives input from MEC (Tamamaki and Nojyo, 1993) and proximal CA1 cells show greater spatial modulation (Henriksen et al., 2010), this suggests that task-specific *Arc* up-regulation reflects the activity of place cells that depend on MEC input, and that changes in environment are encoded by this pathway (Hartzell et al., 2013). Distinctions between *Arc* activity in CA1 and CA3 and between dorsal and ventral hippocampus, which receive distinct topographical inputs from MEC, have also been observed during recent and remote memory recall (Gusev et al., 2005).

Although the function of *Arc* in MEC has not been explicitly addressed, several studies have demonstrated that *Arc* is up-regulated following performance on certain tasks. The requirement for rats to learn or perform a lever-pressing (Kelly and Deadwyler, 2002, 2003) or watermaze task (Gusev et al., 2005; Guzowski et al., 2001), or to simply undergo a swim control (Gusev et al., 2005) up-regulate *Arc* RNA in MEC, although it is unclear precisely which regions of MEC.

Together these data indicate that the expression of *Arc* is a useful indicator of potentially plastic processing, and that although its transcription often corresponds with neural activity, it is likely that the nature of the activity and behavioural history play an important role. The fact that *Arc, Fos* and *zif268* exhibit different patterns of changes under different task demands (Guzowski et al., 2001) is further evidence for this.

2.4.2.3 *From activity to Arc transcription*

If *Arc* is an indicator of synaptic activity, and plays a role in the regulation of plasticity, then what is the process of events that links activity to transcription and eventually to protein function? It has emerged in the last decade that the *Arc* gene is regulated by a number of elements that are directly or indirectly recruited by synaptic activity, and also by the growth factors that also stimulate Arc transcription. Two enhancer elements have been described at 6.5kb and 1.4kb upstream of the *Arc* transcription initiation site (Pintchovski et al., 2009). The distal enhancer region is a highly conserved serum response element that binds the transcription factor, serum response factor (SRF), and is recruited by synaptic activity, forskolin, and

BDNF, while the proximal region also appears to enhance induction by synaptic activity and BDNF (Pintchovski et al., 2009). *Arc* also possesses a promoter with an activity sensor, the synaptic activity response element (SARE), which is a 100 base pair sequence 7kb upstream of the *Arc* transcription initiation site (Kawashima et al., 2009). Activation of the SARE by CREB, MEF and SRF in live animals is sufficient to reproduce the features of Arc transcription (Kawashima et al., 2009). The binding of SRF to particular serum response elements (which are typically possessed by all IEGs (Ramanan et al., 2005)) in the *Arc* promoter is necessary for certain forms of LTD in cerebellar neurons (Smith-Hicks et al., 2010).

The rate of transcription is not only influenced by synaptic activity, but BDNF-induced Arc transcription can also be down-regulated in hippocampal neuron culture and organotypic slices by activation of AMPA receptors (Rao et al., 2006). NBQX, which inhibits AMPA receptors, allows increases in *Arc* RNA, an effect that is occluded by pertussis toxin, which inhibits the G_i-protein that is activated by AMPA receptors (Rao et al., 2006). This suggests a G-protein mediated effect of AMPA receptor activation of Arc transcription, whereas this pathway does not appear to influence translation (Rao et al., 2006). High-intensity stimulus-dependent *Arc* transcription also depends on the activation of NMDA receptors (Lyford et al., 1995; Steward and Worley, 2001b; Coba et al., 2008) whereas LTD-inducing-stimuli-dependent *Arc* transcription can be induced by the mGluR agonist DHPG, which depends on ERK1/2 activity (Wang et al., 2009a).

2.4.2.4 *From activity to Arc translation*

Arc protein is translated from *Arc* RNA, but the extent to which these processes operate in parallel is not clear. *Arc* protein is translated over a sustained period after a learning period and is known to peak in level between 2 and 4 hours, increasing in the soma, synapses and dendrites (Lyford et al., 1995; Link et al., 1995). *Arc* RNA is not simply switched on and continually translated but is subject to nonsense-mediated decay due to the positioning of two exon junction complexes (EJC) on its 3′ UTR (Giorgi et al., 2007), meaning that is often degraded after a single round of translation (Conti and Izaurralde, 2005). In BDNF-stimulated hippocampal neurons, it has been shown that 59% of *Arc* transcripts colocalise with a core component of the EJC complex eIF4AIII, suggesting that they have not been translated prior to reaching the dendrites (Bramham et al., 2010). Depletion of eIF4AIII leads to increases in *Arc* mRNA and protein levels due to lack of nonsense-mediated decay (Giorgi et al., 2007). It has been suggested that Arc is only weakly initiated for translation because of the strong secondary structure of its 5′ UTR (Bramham et al., 2010). Its 5′ UTR also contains an internal ribosomal entry site that may mediate cap-independent translation (Bramham et al., 2010). These features could be important for precisely regulating its translation. There is evidence that the translation of proteins such as *Arc* during LTP consolidation depends on a MAP kinase (MNK1)-dependent mechanism that regulates phosphorylation of the translation factor eIF4E. High frequency stimulation (HFS) stimulates rapid, sustained phosphorylation of eIF4E over a time period similar to the synthesis of *Arc* and inhibition of MNK1 prevents *Arc* protein expression but does not impact on RNA (Panja et al., 2009).

Depending on the type of stimulation, *Arc* translation can depend on NMDA-receptor or mGluR-receptor signalling. The NMDA receptor antagonist APV blocks phosphorylation of

eIF4E and LTP-associated increases in *Arc* (Panja et al., 2009). These effects are not simply due to disrupted transcription process but affect protein translation (Panja et al., 2009). *Arc* protein expression can be further enhanced by the activation of G(s)-coupled dopamine and β-adrenergic receptors (Bloomer et al., 2008). mGluR-LTD dependent translation of *Arc* occurs rapidly in dendrites (Waung et al., 2008) and depends on the phosphorylation of eEF2 (Park et al., 2008), protein phosphatase PP2A and dephosphorylation of the fragile X mental retardation protein (FMRP) (Niere et al., 2012). There is evidence that phosphorylated FMRP acts to suppress dendritic Arc translation as Fmr1 KO neurons exhibit enhanced levels of basal Arc protein in the dendrites but not the soma (Niere et al., 2012).

2.4.2.5 *What cellular processes is Arc necessary for?*

There is an extensive body of work demonstrating that *Arc* is activated by many different stimuli and that it is an indicator of activity, but what are its cellular functions? This question has been approached in two different ways: using Arc knockout animals in which the *Arc* gene has been deleted (e.g. Plath et al., 2006), and through disruption of RNA and translation by use of antisense oligodeoxynucleotides (AS oligos) (e.g. Messaoudi et al., 2007).

Critically, *Arc* is necessary for the protein-synthesis dependent late-phase or consolidation of NMDA-dependent LTP, but not its induction (for a review, see Korb and Finkbeiner (2011)). This was first shown *in vivo* using AS oligos by Guzowski et al. (2001), and has also been shown in a knockout animal in slices (Plath et al., 2006). Infusion of *Arc* AS oligos 2 hours after high frequency stimulation (HFS) led to complete and permanent reversal of LTP and rapid knockdown of induced *Arc* mRNA and protein expression (Guzowski et al., 2001), which suggests that sustained translation of new *Arc* mRNA is necessary for LTP consolidation. At 4 hours the effect was much less (Guzowski et al., 2001). Despite its necessity for LTP, *Arc* does not appear to be essential for AMPA-receptor mediated synaptic transmission (Plath et al., 2006). It is, however, necessary for increases in level of the protein Notch, which is also essential for LTP in postnatal hippocampus (Alberi et al., 2011).

Arc is also, somewhat conversely, essential for certain forms of protein-translation- dependent long-term depression (LTD), including mGluR-dependent LTD (Waung et al., 2008) and cerebellar LTD (Smith-Hicks et al., 2010). It is also involved in homeostatic synaptic scaling (Shepherd et al., 2006; Gao et al., 2010), which is abolished in *Arc* KO animals (Chowdhury et al., 2006).

The mechanism of LTD dependence and homeostatic influences appears to largely involve the endocytosis, trafficking and regulation of receptor subunits of AMPA receptors (Chowdhury et al., 2006; Waung et al., 2008), which occurs through interactions between Arc and dynamin and endophilin and depends on the protein FMRP (Park et al., 2008). Artificial increases in *Arc* using fluorescence-tagged recombinant proteins reduce the size of AMPA-mediated currents in vitro (Rial Verde et al., 2006).

The mechanism by which LTP depends on *Arc* is less clear. Arc interacts with cytoskeletal proteins (Lyford et al., 1995; Fujimoto et al., 2004), leading to the stabilisation of F-actin (Steward and Worley, 2001a; Messaoudi et al., 2007), which facilitates the docking of mRNA to synapses. There is evidence that the stabilisation of F-actin is sufficient to enable LTP consolidation in the absence of *Arc, in vivo* (Messaoudi et al., 2007). The use of *Arc* AS oligos

has shown that regulation of actin dynamics by newly synthesised Arc protein contributes to the maintenance of eIF4E phosphorylation during LTP in the dentate gyrus *in vivo* (Tiron & Bramham, as cited in Bramham et al. (2010)), which enables further translation. These observations suggest that *Arc* positively regulates its protein expression, possibly with the effect of increasing the rate of degradation of *Arc* mRNA as it is translated. This could enhance the specificity of learning to small time periods and specific tasks or environments. The structural effects of *Arc* are also observed in differences in the distribution of spine morphology in *Arc* knockout animals (Peebles et al., 2010).

Recent work has provided a possible explanation for the fact that Arc is stimulated by LTP-inducing stimuli, yet one of its major roles is the internalisation of AMPA receptors. Following neuronal activity, Arc is actually targeted to weak or inactive synapses that previously experienced strong activation, such that AMPA receptors are removed from inactive synapses in neurons that overall are potentiated (Okuno et al., 2012). This form of 'inverse' synaptic tagging (Okuno et al., 2012) may contribute to the specificity of synapses to a learning stimulus. This process depends on CaMKIIβ when it is not bound to calmodulin, acting as a scaffold for Arc at dendritic spines (Okuno et al., 2012).

2.4.2.6 *Behavioural dependencies of Arc*

Given the assumption that LTP is critical for learning, it follows from the previous section that *Arc* also directly impacts on the ability of animals to perform certain tasks. *Arc* has been implicated in many behaviours, including hippocampus-dependent fear conditioning (Czerniawski et al., 2011; Mamiya et al., 2009; Maddox and Schafe, 2011), cortex-dependent inhibitory avoidance training (Holloway and McIntyre, 2011) hippocampus-dependent inhibitory avoidance memory (McIntyre et al., 2005), the sensory deficits of monocular deprivation (McCurry et al., 2010), conditioned taste aversion and delayed object recognition (Plath et al., 2006). Relevant to the entorhinal-hippocampal system, *Arc* is necessary for spatial learning as *Arc* knockouts plateau at a lower level performance during acquisition of the water maze task and are slow to learn the location of a submerged platform when the location is altered (Plath et al., 2006). This may be due to the inability of Arc knockouts to undergo late LTP (Plath et al., 2006). Arc is also necessary for the cellular property of orientation specificity in visual cortex (Wang et al., 2006).

2.5 GENE EXPRESSION, ARC AND THE MEC

In summary, *Arc* is an important candidate molecule for mapping learning in the mammalian brain, positioned as it may be as a "master regulator of protein-synthesis dependent forms of synaptic plasticity" (Bramham et al., 2010). The interactions of *Arc* and its activity-induced expression could be informative regarding activity across MEC circuitry. When combined with an improved understanding of the genomic anatomy of MEC, it could also lead to predictions about whether cells in different layers and regions of MEC contribute distinctly to the modifiable connectivity that likely contributes to the importance of MEC in spatial memory.

3

CHARACTERISING GENE EXPRESSION IN MEC THROUGH ANALYSIS OF ISH DATA

3.1 SUMMARY AND KEY FINDINGS

Gene expression is an important component of neuronal function. In this chapter I outline a resource that I have developed that enables gene expression patterns in the adult mouse brain to be extracted from *in situ hybridisation* data from the Allen Brain Atlas. Gene expression patterns can be extracted at high resolution from any brain region. Here, I have focussed on the analysis of gene expression in the medial entorhinal cortex (MEC) to show that:

1. Gene expression follows patterns across the MEC that are consistent with both anatomical layer boundaries and dorso-ventral gradients in functionally relevant properties.
2. At a conservative estimate, over 100 genes can be categorised as being expressed in a single layer of MEC, but not the others. The majority of these genes are expressed in layer II.
3. A further 2000 genes are expressed at higher levels in one layer than any other. Again, the majority of these genes are higher in layer II.
4. Over 500 genes are expressed in two of the major cell body layers but not in the other. Genes are more likely to be expressed in layers II and III, but not layer V, or layer II and V, but not layer III, but are highly unlikely to be expressed in layers III and V, but not layer II. This suggests that the molecular requirements of cells in layers V and III are usually also requirements of layer II neurons.
5. Genes are most likely to show uniform dorso-ventral expression, but in the superficial layers a large proportion (over 1000) are expressed at higher levels in dorsal than ventral cells, with a much smaller proportion (less than 100) showing higher expression in ventral regions.
6. While dorso-ventral gradients are predominantly observed in the superficial layers, in the deep layers, ventro-dorsal gradients are more commonly observed. A small proportion of these genes actually show opposing gradients across the superficial and deep layers.
7. Relative expression of genes across MEC layers is highly conserved across the medio-lateral extent, which suggests that functional modules with different features are unlikely to be observed along this axis.

3.2 INTRODUCTION

Gene expression is an important determinant of neuronal identity and function. Gene expression patterns reveal not only cellular diversity, but correspond with structural and functional delineations. They may therefore reveal regional differences in the types of computations being performed and may underlie cell-type specific and layer-specific vulnerability to disease. As such, characterising gene expression patterns may aid in our understanding of neural circuits, and enable us to manipulate them specifically. Until recently, such analyses were limited by both the experimental and computational tools necessary to study patterns of large numbers of genes but more recently, such problems are becoming increasingly tractable.

The molecular organisation of the medial entorhinal cortex (MEC) has not been characterised. The MEC is a component of the spatial navigation circuit and is believed to be critical to the encoding and storage of spatial representations, a characteristic that likely depends on the expression of certain genes. There is anatomical and physiological evidence that layers and specific cell types within the MEC have specific functions (see Section 2.3.2), but we currently have limited methods for independently controlling them to thoroughly test hypotheses. What makes this circuit particularly interesting is the extent of topography; dorso-ventral gradients in physiological (Hafting et al., 2005) and cellular characteristics (Garden et al., 2008) are observed in MEC that map anatomically to septo-temporal gradients in the hippocampus (Dolorfo and Amaral, 1998b; van Groen, 2001). It therefore seems prudent to attempt to map the molecular topography of MEC, using gene expression patterns, to connectivity, regional function, and ultimately behaviour.

Elucidating the genetic profile of the MEC has two further potential benefits. First, the anatomy of the region is difficult to characterise and has multiple definitions depending on whether cytoarchitecture, connectivity or molecular labelling is used to define it (reviewed by Canto et al. (2008)). For example, in mouse, there is no universal agreement as to the regional identity of the cells that occupy the region dorsal to MEC (c.f Burgalossi et al. (2011); Boccara et al. (2010)). Second, multiple diseases with a demonstrable genetic component, including Alzheimer's disease (Bertram and Tanzi, 2008; Braak and Braak, 1991) and other neurodegenerative diseases (Braak et al., 2000), schizophrenia (Arnold, 2000), autism and epilepsy (see Section 2.2.6.2), are all linked to pathology in the MEC, but there is little to directly connect such genes to the MEC. One possible route to understanding the specificity of layers to cognitive processes and disease is to study the distribution of disease-related genes in vulnerable regions.

Large-scale investigations into gene expression patterns have previously been carried out in the neocortex (Belgard et al., 2011) and sub-cortical structures such as the striatum (Doyle et al., 2008) using techniques that include translational profiling, microarray or RNA-seq analysis of dissected or cell-specific labelled tissue (see Section 2.1.1 for an overview). RNA-seq analysis has revealed that thousands of protein-coding genes, as well as non-coding RNA loci, are differentially expressed across layers of somatosensory cortex (Belgard et al., 2011). RNA-seq has the strength of being a highly quantitative method with a large dynamic range that is not limited by probe-based methods. However, this type of analysis relies on pre-existing knowledge of anatomy or use of transgenic lines in which specific cell types are

labelled, and is therefore not highly sensitive to identifying novel patterns of gene expression that are potentially related to anatomy.

A different approach involves using ISH data to extract gene expression patterns. ISH methods can be naive to anatomy, but quantification is a difficult problem because ISH results depend strongly on probe-binding efficacy, which is unlikely to be equivalent across genes. The Allen Institute has provided a tool that overcomes as many of these limitations as possible and has been used to investigate gene expression patterns in the somatosensory cortex (Lein et al., 2007), hippocampus (Thompson et al., 2008), and cerebellum (Kirsch et al., 2012). Since much of the anatomy of the MEC is not understood genetically, and therefore cell types and layers cannot be specifically targeted, this approach could provide an appropriate means of characterising gene expression patterns. In the next section I outline the methods used to create the ABA, and identify their limitations for studying gene expression in MEC. I also outline the benefits of using an alternative method for this particular task.

3.2.1 *The Allen Brain Atlas data*

The Allen Brain Atlas is a high-quality comprehensive online repository of *in situ hybridisation* images (Lein et al., 2007). It contains over 26000 samples of 20,382 genes. To make the data as useful as possible for the research community, the process of ISH and collection of image data were designed to have a high signal-to-noise ratio and be as similar as possible across genes. As such, all experiments used mice of the same species (C57Bl/6J), age (56 weeks) and gender (male) to minimise the effects of inter-animal variability. In addition, all ISH images were produced using identical ISH protocols and imaging conditions and have been through a quality control procedure (see ISH Platform Controls pdf). Nevertheless, the nature of this data presents several challenges. First, it is difficult to obtain a method for ISH that provides a large dynamic, measurable range. Second, it is difficult to effectively assess the quality of such a large range of experiments. The third challenge is organising vast amounts of data from highly variable brain structures into a system that can be of use to all. I will address each of these challenges here and highlight how my work intends to overcome the latter two.

The raw data: *In situ hybridisation* is a process by which an RNA probe is applied to freshly prepared tissue where, given appropriate conditions, it hybridises to RNA strands. To develop the ABA, the Allen Institute tackled the first challenge by identifying, testing and optimising probes based on maximum signal detection and specificity, and all were applied to tissue in an identical manner. Nevertheless, probe design impacts on binding efficacy, particularly in terms of probe length and GC content. For the ABA RNA riboprobes were designed using a semi-automated process. DNA templates were acquired either from cDNA clones (10000), genomic tail DNA or pooled cDNA from mouse brain (9000). cDNA clones were used as direct templates, whereas probe sequences for cDNA templates were taken from the region within 3000bp of the 3' end. The 1000 probes generated from tail DNA were designed within single exons with a minimum length of 400 bp. All probes were designed to be between 300 and 1200 nucleotides, ideally greater than 600 and with a GC content between 42-62%. Probe sequences were tested for homology to minimise cross-reactivity within gene families.

Although these features were carefully controlled, differences in the length and GC content of riboprobes could contribute to differences in binding efficacy and tissue penetration in an unpredictable manner. As such, the use of ISH to compare absolute levels of gene expression between genes is limited.

To enable detection of RNA, the method of ISH used a non-isotropic colorimetric assay to stain RNA-positive cell somata a blue/purple colour (See Fig 3.1(A)). First an RNA probe was hybridised to RNA in the tissue. This probe was conjugated to a DIG label that could be subsequently detected by an anti-DIG antibody. Two subsequent amplification steps were used to ensure optimal signal to noise, which was followed by a final alkaline phosphatase stage in which a blue/purple precipitate formed (with the use of NBT/BCIP) on the tissue at locations to indicate the presence of RNA (ABA Data Production Processes). The application of this method across thousands of genes means that, as for microarray data, gene expression can be simultaneously compared across genes, to some degree.

For the ABA, ISH was performed on coronal or sagittal 25 μm thick sections, sampled at 25 or 200 μm, respectively, and images were obtained at high-resolution using bright-field microscopy. Corresponding expression images were generated for the ABA by automated detection of stained cells in ISH images, followed by the extraction of the expression level and size characteristics of each expressing cell. In this way, each ISH image is accompanied by a greyscale image representing the absolute level of expression of the gene within cells, which can be colour-coded for ease of viewing (See Fig 3.1(B)). The quantification of ISH data was performed by extracting pixel intensity data from the expression images because, in theory, they contain only gene expression relevant data and do not include anatomical features. However, while the cell detection algorithm used for the ABA was typically very accurate, there are examples in expression masks where the pial surface and debris have been detected as cells (see Fig. 3.1(C)), a problem that is very difficult to automatically detect and that compromises some of the data. In the work that I describe in this chapter, I have used classification methods to identify individual ISH images that potentially show this error.

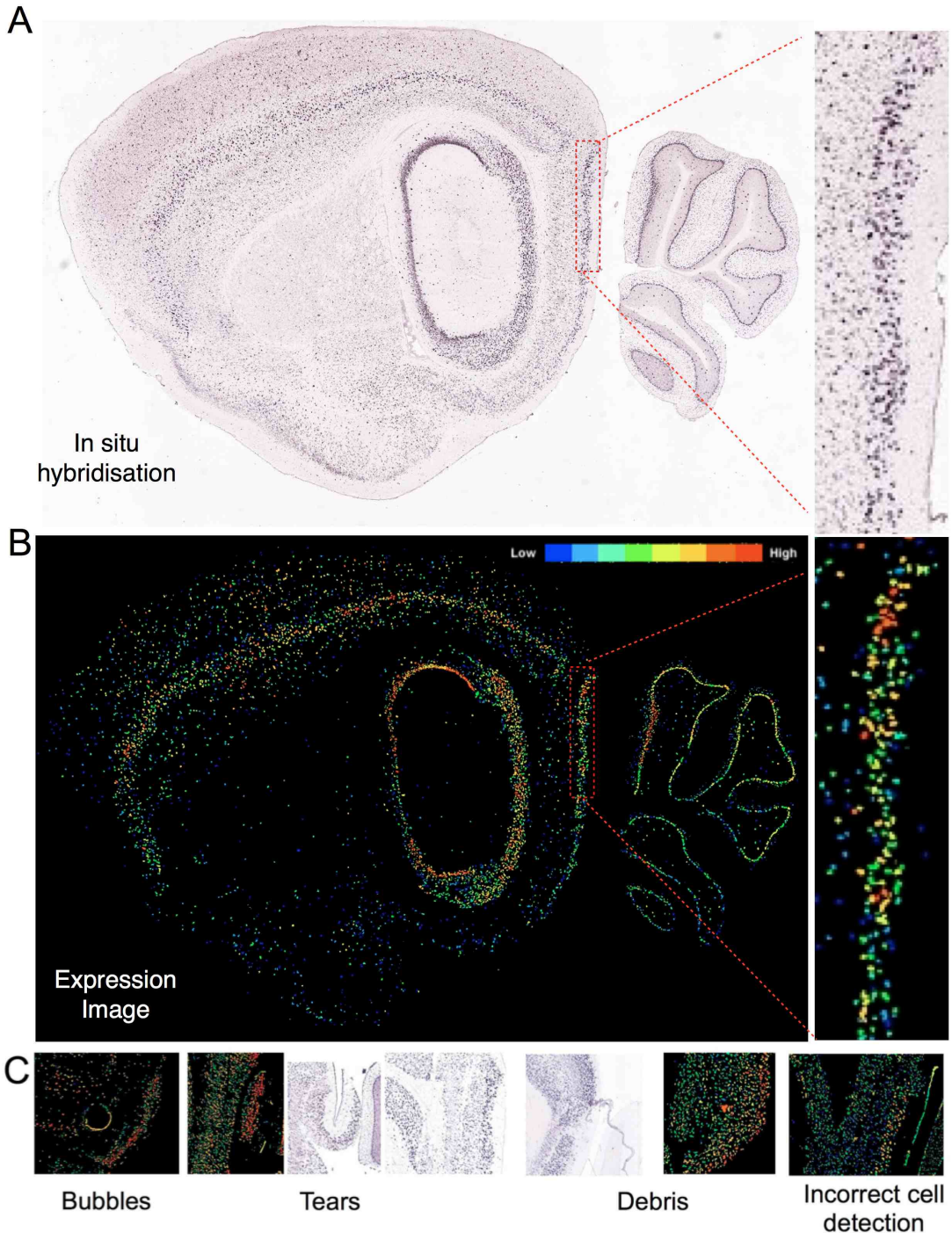


Figure 3.1: *Example of Allen Brain Atlas data* (A) *In situ hybridisation* image from the Allen Brain Atlas, showing RNA expression of the *Hcn1* gene, indicated using a non-isotropic colorimetric assay. A region in the dorsal MEC layer II is shown at higher magnification to emphasise that individual cells can be identified. (B) Automated expression mask of the ISH image for Hcn1. The mask shows all cells with a threshold intensity of staining. Differences in intensity between cells are colour-coded with red reflecting high expression, and blue low expression. The range of this scale is the same for all gene images, so red cells correspond to the highest staining observed for the RNA of any gene. (C) Example images showing different types of error that can affect data extraction, including bubbles, tears, debris and poor cell detection.

Integrating the data into a brain-wide genome-wide resource The ABA uses a three-dimensional registered volumetric representation of gene expression to provide a single resource that contains all data that can be mined and quantified. To generate this volumetric representation, the Allen Institute had to map information from all their 2D ISH images into 3D space (Fig. 3.2). There are several reasons why this is not a trivial problem: (1) there is variability between the different mice used for RNA expression measurements, which means that brain structures do not occupy exactly the same 3D coordinates, (2) when brains are sectioned and applied to glass microscope slides, the shape, alignment and structure of the section is not entirely or systematically preserved, (3) the efficacy of the hybridisation varies between genes, which means that anatomical structures can be delineated to different extents across genes.

The Allen Institute minimised these difficulties by first generating an Allen Mouse Reference Atlas tool, which is an atlas of the mouse brain generated specifically for their dataset and for ISH analysis. Their next challenge for the 3D representation was to identify the true positioning of each section in 3D by identifying the nearest reference atlas section and effectively aligning the section to the atlas (Fig. 3.2(i)). Once an image had been mapped to an atlas section, it was deformed using non-linear warping to match the reference image (Fig. 3.2(ii)). This transformation was applied to the corresponding expression image, from which cell expression data could then be extracted (Fig. 3.2(iii)).

Although this process appears to have been extremely successful, it is difficult to quantify this, and since accurate data analysis relies on accurate registration, potential errors in registration should be taken in account. More recently, researchers investigating the organisation of cell-specific genes across the brain used further 2D image registration on coronal images to ensure correct alignment (Ko et al., 2013). Since the accuracy of 2D image registration can be measured with a similarity metric, this method not only adds to the accuracy of extracting gene enrichment, but also the level of confidence that can be placed in the results. This is therefore a useful alternative to 3D registration if only a small region is being studied.

The strength of the ABA is that it contains a 3D representation of every single gene, such that the gene expression level of any gene can be probed at any voxel in the brain. The atlas also incorporates details of the features of each expressing cell within a voxel, the positioning of the section and cell within the section, and the density of gene expression at each voxel relative to its neighbours. The BrainExplorer tool (Ng et al., 2007b) allows users to visualise gene expression in any brain region in 3D, for example the retro-hippocampal formation (RHP), which includes the MEC (Fig. 3.2(iv)). The colour and size of each 3D voxel represents the summed expression across all the cells in that 200x200x200μm^3 region (Fig. 3.2(v)). This region is relatively large when considering that the main cell body layers of MEC are only 200μm in width, and some anatomical variations occupy an even narrower window. If one takes into account the possibility that there may be registration errors, this voxel size is a clear limitation. An analysis tailored to the MEC would therefore require data extraction to be performed at a higher resolution.

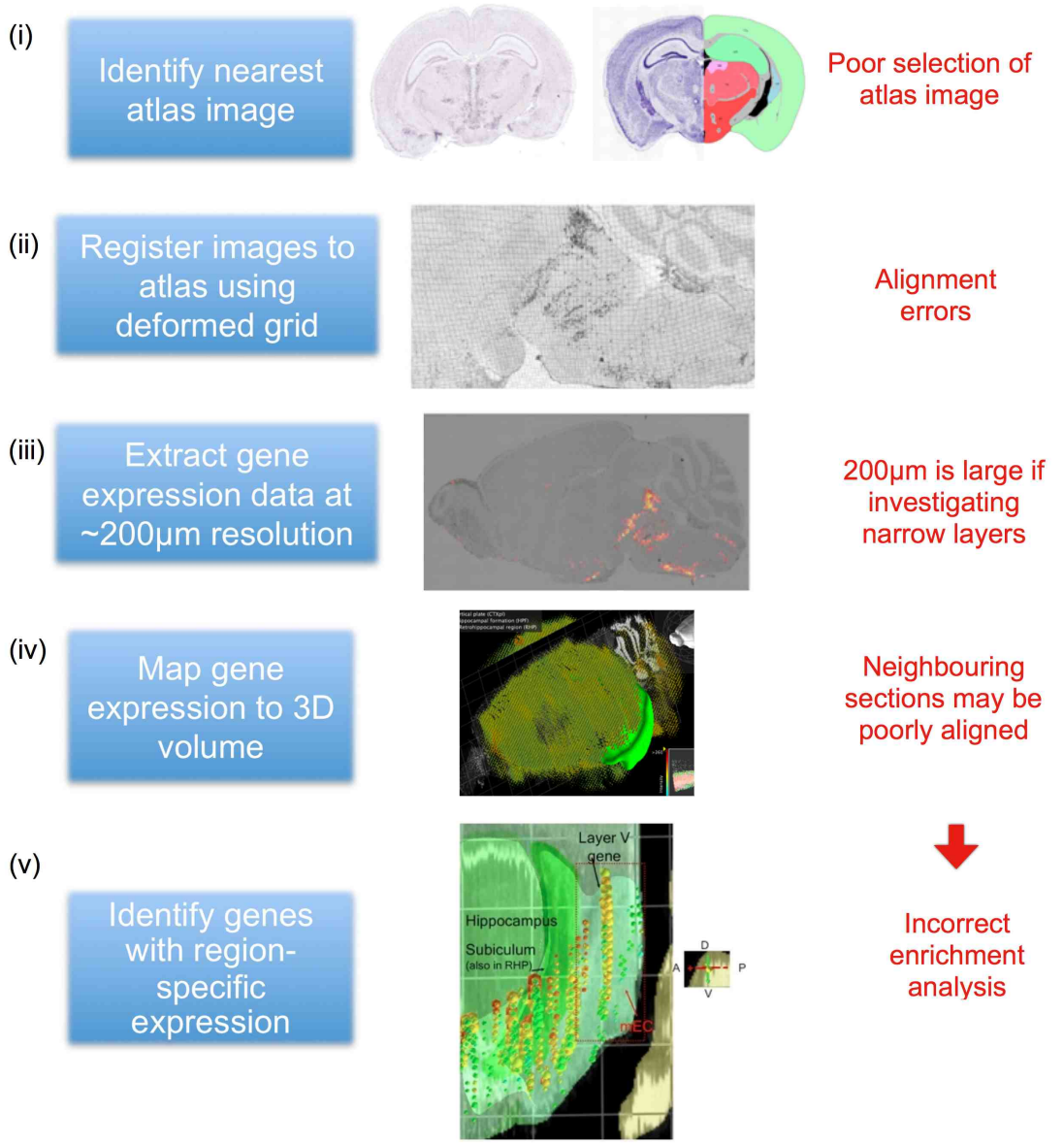


Figure 3.2: *Generation of the Allen Brain Atlas 3D volumetric dataset* Schematic showing the outline of the procedure used to generate the 3D volumetric dataset used by the ABA. For each section (left) (i) the nearest atlas image is identified (right), and (ii) the image is deformed to match the atlas image using registration algorithms. (iii) Pixel intensities are averaged across regions to give gene expression intensity at a 200 µm resolution (colour-coded blobs). (iv) All images for each gene are consolidated to give a 3D representation of gene expression. Brain regions can be identified in 3D, such as the RHP region shown here in green, shown here using the Brain Explorer II.(v) Region-specific expression can then be extracted across all genes. Errors associated with each stage are shown in red. Images are taken from the Informatics pipeline guide.

Using the ABA resource The ABA contains ISH data, acquired by highly similar means, across the entire brain, making it an ideal resource for computational mining and user-tailored queries. The Allen Institute have also developed tools that allow users to identify high gene expression, gene enrichment and differential expression across the entire brain. The tools that are of particular relevance for identifying all enriched genes in a particular

region include the Fine Structure Annotation (FSA), the Anatomical Gene Expression Atlas (AGEA), the NeuroBlast function, and a related tool, the AllenMiner. The FSA provides a list of 50 manually verified genes that show specific patterns of enrichment in coronal sections within a particular area, often containing more than one structure. The AGEA Gene Finder function (Ng et al., 2009) is an extremely powerful tool that allows the user to specify the coordinates of a point of interest then calculates the genes with the most highly correlated gene expression to that point. It therefore provides better resolution of gene enrichment than the FSA. Figure 3.3 shows two spatially segregated areas of correlation; one within the dorsal MEC, and one in ventral MEC to show the difference in correlation profile. Note how both points correlate relatively highly with other areas of MEC and some areas of cortex but show lower correlation with the remainder of the brain. The only problem with such a tool is that the results have not been manually verified, so due to registration errors and image flaws, it yields false positives.

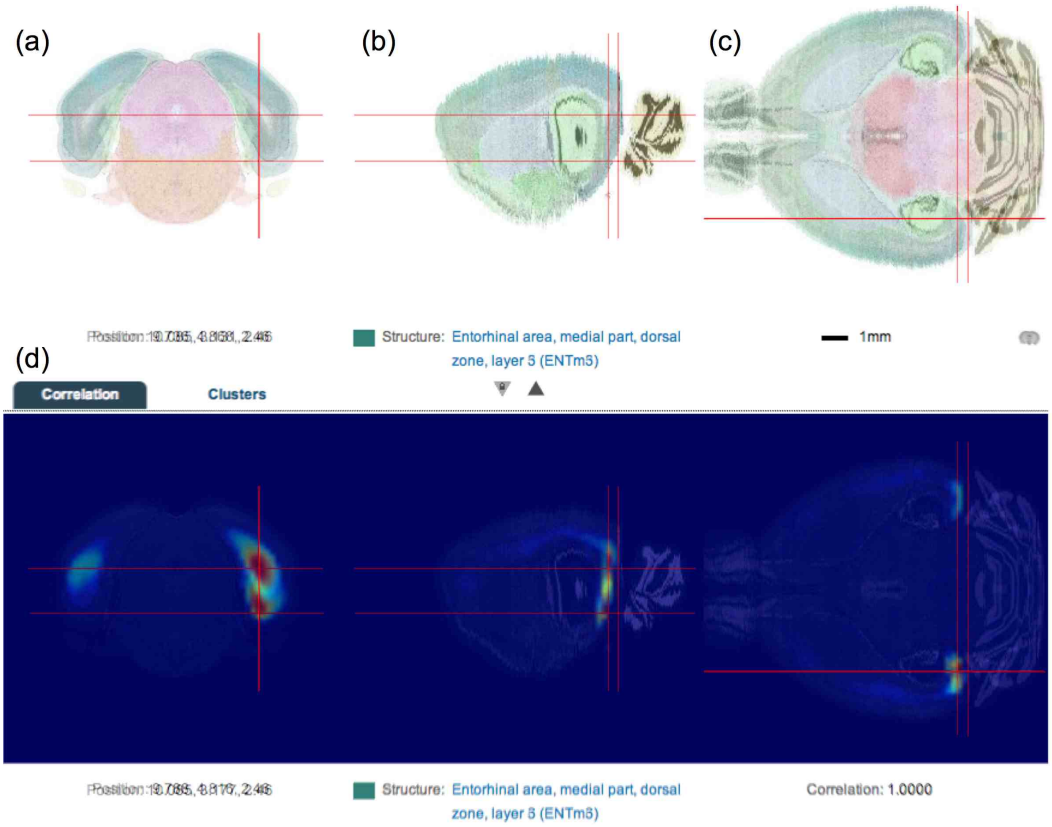


Figure 3.3: *AGEA tool for selecting model genes* Overlaid screenshots of the AGEA tool showing the seed points set at dorsal and ventral MEC locations on the reference atlas in (a) Coronal view, (b) Sagittal view, (c) Horizontal view. (d) A correlational analysis shows that separate regions of high correlation are observed. This difference can only be observed when looking in the top 15% of correlations.

Once genes of interest have been identified, the NeuroBlast function uses the correlations between genes to identify genes with similar patterns to a gene of interest. As such, the ABA tools provide a powerful mechanism of identifying enrichment. The AllenMiner is a tool that was developed by an independent group (Davis and Eddy, 2009) that provides an even more flexible means of extracting data using the ABA API. Most significantly for a project

investigating the MEC, it enables users to define their own regions of interest, unlike the FSA and NeuroBlast, which are limited to brain structures, and unlike the AGEA, which is limited to single coordinates.

The validity of this resource: ISH experiments performed for the ABA have been subject to several types of validation. The top 1000 accessed genes have been directly verified, where possible, with resources from the literature and other online databases. In addition, investigations into gene enrichment in the mouse neocortex using this approach (Lein et al., 2007), produce highly consistent results to RNA-seq data (Belgard et al., 2011). Of particular interest, this atlas has been used to show that the hippocampus can be genomically differentiated into subregions that parallel anatomical changes (Thompson et al., 2008). Gene expression gradients in area CA3 map on to an anatomical axis from the septal hippocampal region (dorso-medial) to the temporal region (ventral) (Fig. 3.4). Whether this translates into differences across the dorso-ventral gradient in MEC in terms of gene expression is currently unknown.

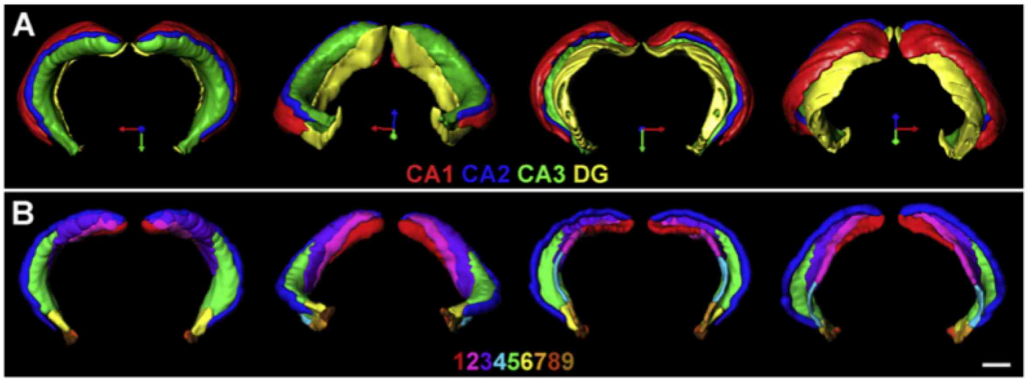


Figure 3.4: *Genetically defined hippocampal sub-divisions:* Thompson et al. (2008) used non-negative matrix factorisation (NMF) to segment the hippocampus according to gene expression patterns. (A) 3D Schematic showing the 4 distinct genomic groups that correspond to the CA1, CA2, CA3 and DG regions, from 4 different angles. (B) 3D schematic showing the 9 genomic modes that the CA3 region can be divided up into, shown from different angles. These regions follow a septo-temporal gradient, showing that different combinations of genes are expressed along this axis. Adapted from Thompson et al. (2008) Fig. 3.

In summary, the ABA have collated an impressive resource of data on the expression of the majority of genes in the brain. In addition, Allen Institute tools provide an ideal means of identifying a number of genes that show selective enrichment in brain regions such as the MEC. However, it is the raw data images themselves that, re-aligned to increase accuracy and improve confidence, I believe will provide an ideal resource for identifying all the genes with patterned expression, down to a resolution of 10 µm in the MEC. This will enable me to identify individual and combinations of genes that define precise laminar organisation, as well as elucidate novel patterns of expression. This type of analysis should enable the mapping of gene expression to functionally relevant properties. At a later stage it could

enable subsequent identification of regulatory mechanisms that mediate and help maintain regional functions within MEC.

3.2.2 *Aims*

- To develop an accurate and reliable method for extracting the expression patterns within the MEC of all genes within the Allen Brain Atlas ISH database. Accurate re-registration of all 2D images containing the MEC, followed by extraction of pixel intensity data, is likely to be the most effective tool. This technique should be applicable to most, if not all, brain regions.
- To use this tool to determine whether gene expression is consistent with the major anatomical layers and borders of MEC. Across neocortex, research has shown that genes are differentially expressed across layers and in the hippocampus regions can be genetically differentiated, suggesting that borders and layers of the MEC will also be genetically defined.
- To determine if gene expression is enriched to differing degrees across different layers. The layers of the MEC contain different cell types and their major projections terminate in different regions, which are characteristics that lead to the prediction that they exhibit different genetic profiles.
- To investigate gradients in the expression of genes and how these vary across layers. This is particularly relevant for questioning whether there are links between septo-temporal gradients in gene expression in hippocampus and dorso-ventral gradients in MEC. Anatomical projections between these regions are topographically organised.
- To identify lists of genes that show robust enrichment in a particular layer of MEC compared with other MEC layers, or that show a dorso-ventral gradient in expression.

3.3 METHODS

This section outlines the methods I have used to establish gene enrichment in the MEC using primary data from the Allen Brain Atlas database. I approached this problem from two complementary angles.

First, to begin to identify enriched genes, I used tools generated by the Allen Institute that extract gene enrichment information from a three-dimensional volumetric dataset. This dataset includes expression information for all genes at 200µm voxel resolution throughout the entire brain, and therefore enables the identification of genes that are specifically expressed in regions that are sufficiently sampled at this resolution. The tools necessary for this approach are readily available and it's possible to quickly identify enriched genes within the MEC. However, the 3D dataset is limited by spatial resolution, alignment errors and errors in generation of the expression masks, thereby limiting the accuracy with which it can identify true enrichment. There is no mechanism of identifying potential errors, so every image must be visually assessed before a gene can be confidently classified.

Second, to provide a more quantitative, high resolution analysis of gene expression patterns, I developed a method to extract data directly from 2D images of sections containing the MEC. The success of this method can be tested with automated measures, and it can be optimised to detect as many enriched genes as possible by using visually assessed genes that were identified using the previous method as training data. However, some limitations do remain with this method because the extraction of valid data relies on a registration process that is not completely verifiable. The procedure for these methods is outlined below and summarised in Fig. 3.5:

1. Approach 1: Identification of enriched genes using pre-existing tools (Fig. 3.5 (blue))
    a) *Use of Allen Brain Atlas tools* to identify a small number of genes with specific patterns of enrichment in MEC
    b) *Visual assessment* of compiled gene lists to identify 'model' genes enriched in each MEC layer or with a dorso-ventral gradient in expression
    c) *Use of the NeuroBlast tool* to find all genes with similar expression patterns to 'model' genes.
    d) *Visual assessment* of NeuroBlast lists to give full lists of genes for each layer or gradient
2. Approach 2: Identification of enriched genes using a tool I have developed that extracts data from 2D images (Fig. 3.5 (green))
    a) *Two dimensional image registration* of all sagittal sections containing the MEC
    b) *Extraction of data* from registered images
    c) *Optimisation of data classification* by using previously visually assessed images
    d) *Final compilation of gene enrichment lists*

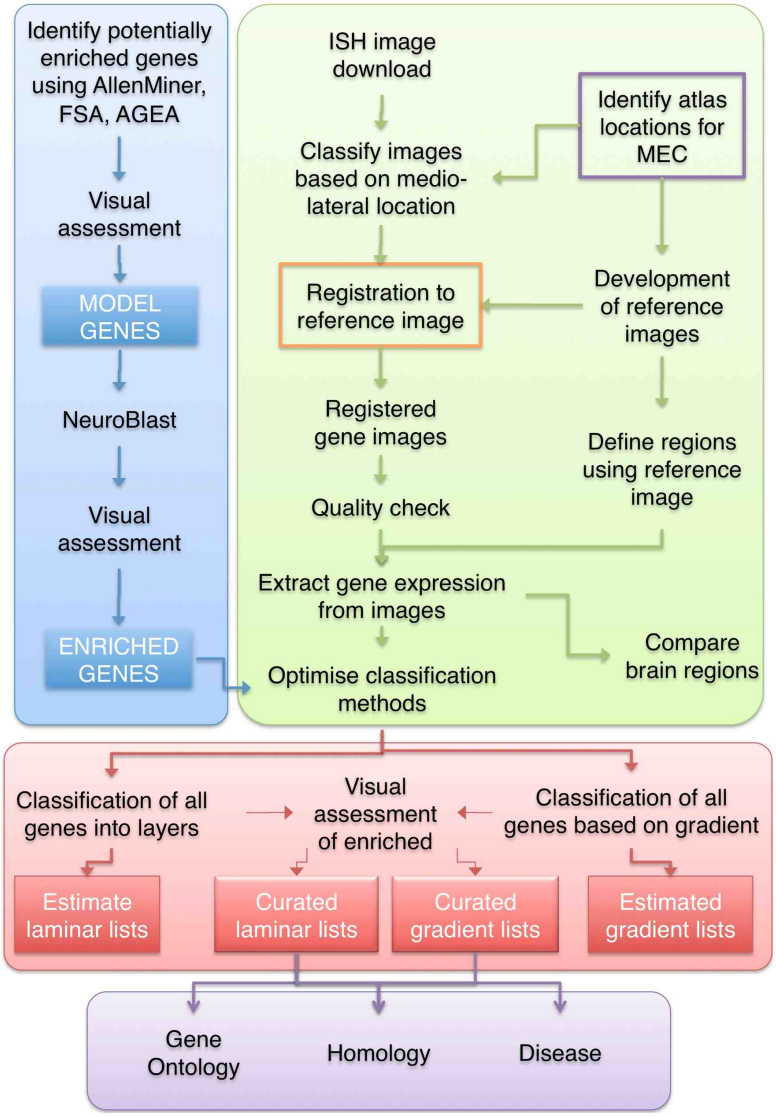


Figure 3.5: *The pipeline for extracting gene expression data from 2D images* This workflow shows how the two methods: (1) use of ABA tools (blue) and (2) data extraction from 2D images (green)- are combined to identify enriched genes. The stage outline in purple is shown in Fig. 3.6 and the stage in orange is shown in Fig. 3.13. The eventual aim for this chapter is to generate complete lists of all enriched genes that include manually validated, and estimated lists (where lists are too large to validate manually) (red). In the next chapter I will address whether patterned genes are enriched for particular gene functions (purple).

3.3.1 *Identifying potentially enriched genes using Allen Brain Atlas tools*

The most powerful tool provided by the Allen Brain Atlas to generate lists of region-specific genes is the NeuroBlast tool, which uses the 3D registered dataset to find all genes with similar expression patterns to the input gene. To generate a list of model genes for each layer that I could use as input to NeuroBlast, I investigated several unbiased methods. These included the Fine Structure Annotation (FSA), the Anatomical Gene Expression Atlas (AGEA) Gene

Finder and the AllenMiner (See Fig.3.5)[1]. These methods all extract data from the 3D volumetric dataset described but their methods are different, thereby providing complementary approaches. For all methods I made use of sagittal sections (all taken from the left hemisphere) because the MEC can be rendered most unambiguously from this angle. I then visually assessed the raw ISH images of all genes identified by these methods using the criteria defined in Section 3.3.2.1. This enabled me to identify specifically expressed genes across all layers and with various gradients.

3.3.1.1 *Use of the Fine Structure Annotation*

I used the FSA to identify genes with specific enrichment in the MEC. I searched within a region termed the entorhinal area (the smallest region that could be chosen for studying the MEC), which identified a number of specific genes that show enriched MEC expression compared with the brain and therefore may be more interesting to MEC function. The limitation of this tool was that neither the MEC nor its layers could be individually selected, so although there were no false positives due to image or alignment errors, genes had to be visually checked to ensure they were patterned in MEC, and this method revealed few genes.

3.3.1.2 *Use of the AGEA*

The AGEA tool identifies coronal images of genes that are highly correlated with particular 3D voxels, so I used this to identify genes enriched in voxels that fall within the MEC region. As sagittal results were necessary for my purposes, I identified gene results in the AGEA search and visually assessed the equivalent results from the sagittal plane. I attempted to identify dorso-ventral genes by using seed locations within the dorsal and ventral MEC, as shown in Fig. 3.3 but although genes could be identified for each region, it was difficult to explicitly identify genes higher in the dorsal region than the ventral region of MEC. Also, as predicted, a significant proportion of genes were not patterned in the region of interest. However, this procedure allowed me to identify further 'model' genes than the FSA had identified.

3.3.1.3 *Use of the AllenMiner tool*

The previous methods were not particularly sensitive for identifying genes with gradients. Fortunately, a tool developed by Davis and Eddy (2009), the AllenMiner, is ideal for this type of task. It essentially mines the 3D database allowing users to manually select combinations of voxels from which expression data can be extracted. Use of the tool required several steps: (1) voxel selection from the 3D atlas to select out layers and dorso-ventral regions, (2) specificity and enrichment analysis to identify specific genes, (3) gradient analysis on dorso-ventral regions. I decided not to use this tool to identify layer-specific genes because of the difficulty in precisely defining layers using voxels of 200x200x200 μm^3.

Defining dorso-ventral subregions within the MEC Reference atlas images containing the MEC (Figure 3.6(B)) were used to make MEC partitions in 3D space. I used three sagittal sections

1 These tools were available and most appropriate for use in June 2011 when this section of work was completed. In November 2011, significant changes were made to the ABA website, including to some of these tools.

(ML2-ML4) and generated partitions independently for each so that medio-lateral differences could potentially be investigated. I identified the boundaries of the MEC and wrote a Python script to convert 2D coordinates from the reference atlas into 3D coordinates used by the AllenMiner and BrainExplorer. The most dorsal, ventral, anterior and posterior 2D coordinates were identified from either the AGEA (Ng et al., 2009) coordinates (in Allen Reference terms) or Allen Reference Atlas coordinates, which weren't precisely in accordance.

The Allenminer tool was used to separate the MEC partition into four DV partitions so that expression for each gene within each partition could be summed to calculate a gradient. Partitions were made perpendicular to the DV axis (Fig. 3.6(C)). The 3D locations of the partitions were validated by modelling the user-defined partitions, using the retrohippocampal region as a landmark (Fig. 3.6(D)), using the software PyMOL, which enables 3D rendering (See Figure 3.6(E)). PyMol models were then compared to the BrainExplorer representations (Lau et al., 2008) and 2D images of a number of highly specific genes. The cutting tool in BrainExplorer allowed me to specifically isolate voxels expressing a gene present at the desired reference atlas position so that I could check that the coordinates matched.

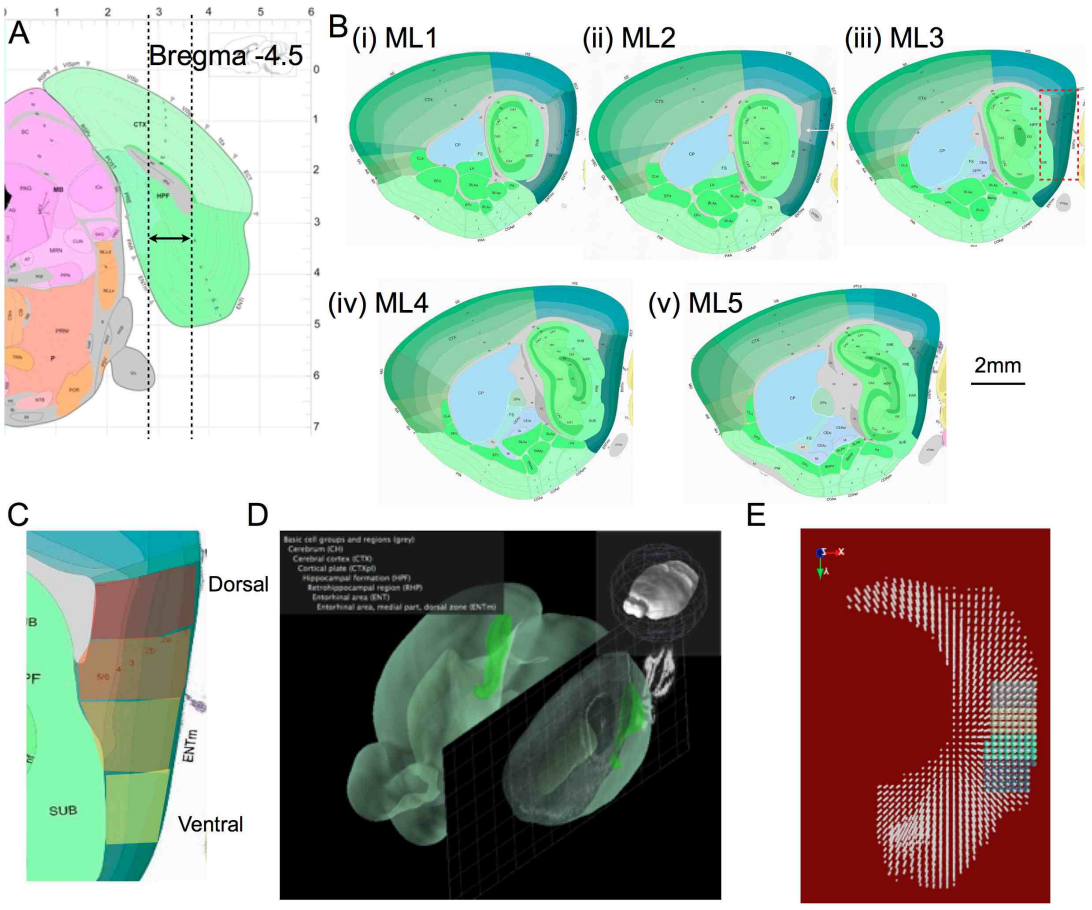


Figure 3.6: *Reference atlas images used to define the MEC location* (A) Reference atlas section (taken from the ABA reference atlas) in the coronal plane. Dotted lines indicate the medio-lateral extent of MEC. (B) Reference atlas sections (taken from the ABA reference atlas) in the sagittal plane indicating the 5 reference regions containing MEC or its bordering regions. The red box outlines a region shown at higher magnification in C. In ML1 and ML2 images, this atlas suggests that lateral EC lies dorsal to MEC, but this is not consistent with the atlas of Franklin and Paxinos, which shows that caudal entorhinal area (which corresponds to MEC), extends the full dorso-ventral extent of the entorhinal area even at very lateral locations. (C) Dorso-ventral partitions throughout the MEC. The partitions used to detect gradients using the AllenMiner tool are shown overlaid on a reference atlas section. (D) Screenshot showing the 3D MEC in the Allen Institute's Brain Explorer II. (E) Three-dimensional PyMol (The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC) rendering of the MEC partition (coloured) within the retrohippocampal region (white) showing voxels corresponding to the four partitions from dorsal to ventral regions, spanning the whole anterior-posterior and medio-lateral extent. These divisions were used to query the AllenMiner for gradients in expression.

Identifying gradients I used the Allenminer tool to calculate gradients along the DV axis based on a consistent reduction or increase in expression from dorsal to ventral MEC. This relied on calculations of specificity and enrichment for each gene (Davis and Eddy, 2009). Specificity is calculated as the proportion of the expression of the gene that is found in the region of interest (ROI) (see Eqn:3.2). Enrichment incorporates the volume occupied by the ROI (see Eqn. 3.3). Thus, a gene with the same specificity in the MEC as in the layer II region has a much higher enrichment in MEC layer II. I used the AllenMiner tool (Davis and Eddy,

2009) to calculate, for each gene, the expression level in each ROI along the DV gradient, then its specificity and enrichment. This data could then be used to calculate gradients (See Eqn: 3.4). A regional patterning score was also calculated to represent the uniformity of expression across the four partitions (See Eqn: 3.5, where i is the inner partition, n is the total number of partitions.).

$$\text{Specificity}_{roi} = \frac{\text{Expression}_{roi}}{\text{Expression}_{brain}} \tag{3.1}$$

$$\text{Enrichment}_{roi} = \frac{\text{Specificity}_{roi}.\text{Size}_{roi}}{\text{Size}_{brain}} \tag{3.2}$$

$$\text{Gradient}_{roi} = \frac{\sum_{i=2}^{n} \text{Enrichment}_{roi_i} - \text{Enrichment}_{roi_{i-1}}}{\sum_{i=2}^{n} \|\text{Enrichment}_{roi_i} - \text{Enrichment}_{roi_{i-1}}\|} \tag{3.3}$$

$$\text{RPS}_{roi} = \sum_{i=2}^{n} \text{Specificity}_{roi_i}.\log_2(\text{Enrichment}_{roi_i}) \tag{3.4}$$

$$\tag{3.5}$$

3.3.2 *Generating lists of enriched genes*

These three tools generated a considerable number of potentially specifically expressed genes in MEC. I visually checked all images to verify the enrichment because the datasets are sensitive to misalignment and image quality errors. The AllenMiner generated a large number of genes but due to a very high rate of false positives, I decided to check just the most high scoring, and use the NeuroBlast from this point.

3.3.2.1 *Criteria for visual assessment and classification of 'model' genes*

The ISH (data type 1) and expression image data (data type 2) were visually assessed and categorised according to localisation of strong labelling within layers and the presence of gradients. The types of patterns that reflect enrichment in a particular layer are shown in Fig. 3.7. The gene categories that genes were assigned to are shown in Table 3.1. They were not unique, but were designed to capture the predominant patterns of expression. For example, 'LII & III' genes could also be classified in 'LII' if they showed low expression in layer V, but also much higher expression in layer II than in layer III. Genes were not only assigned to a layer, but were also rated based on my confidence in the classification: 3: Specific expression with high absolute level; 2: Very clear specific expression, 1:Clear specific expression. Only genes with high (2 or 3) ratings for each category were used to generate the model genes lists that were input to NeuroBlast.

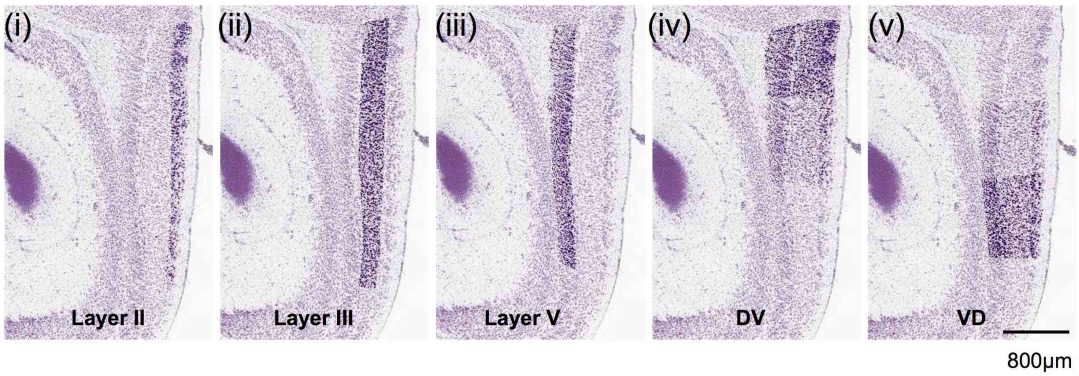


Figure 3.7: *Examples of patterns of expression for 'model' genes.* Images of Nissl-stained sections in which specific layers (i-iii) or regions of MEC (iv-v) have been artificially enhanced to show the types of staining patterns that would reflect gene expression enrichment.

Category	Description
Layer II	Present in layer II only or higher in layer II than any other
Layer II & III	Present in layers II and III
Layer III	Present in layer III only or higher in layer III than any other
Layer V	Present in layer V only or higher in layer V than any other
Uniform	Genes that show staining in all layers
None	Genes that show no staining in the expression mask
DV	All genes with any form of DV gradient
DV layer II	All genes with a DV gradient in layer II
VD	All genes with any form of VD gradient
No gradient	Genes showing no gradient

Table 3.1

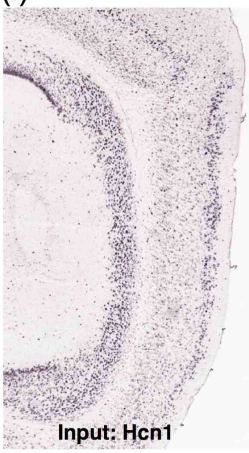
3.3.2.2 *Using the NeuroBlast tool to find genes with similar expression patterns to model genes*

Once all genes had been checked, I took exemplary model genes (~ 20 in each list) and used the NeuroBlast feature from the ABA (Ng et al., 2007b) to find genes with similar expression patterns within the retrohippocampal region (the search could not be narrowed to the MEC). The tool detects correlations between genes using the ratio of expression energy scores across sagittal sections. Expression energy is a measure that depends on the expression intensity of a gene at a particular voxel in 3D and its expression in surrounding voxels, and is therefore sensitive to enrichment, rather than just high expression. Fig. 3.8 shows example NeuroBlast correlation results for the gene Hcn1.

I used a custom-written Python script to run the NeuroBlast function for all the model genes within each model gene list, extract the xml results and generate a list of all genes with

correlated expression to at least one model gene. The correlation score between all genes was recorded such that I could subsequently choose a threshold correlation and assess by eye all genes exceeding that threshold, the assumption being that high correlations are more likely to yield valid gene results. I set thresholds at 0.5, 0.55, 0.6, 0.7, 0.8 and 0.9 and calculated for each threshold the proportion of genes detected by NeuroBlast that matched the input model category according to my visual assessment.

(i) (ii) (iii) (iv)



Gene Name	Image Series	Correlation Score
Kpnb1	70429285	0.786177
Mm.26272	75692997	0.746052
Rhog	68745114	0.707444
Ppp1r2	69817217	0.695586
Arid5b	71283598	0.6907
Eif1	67978400	0.686113
Gpr161	70562045	0.673436
Abcf2	75695665	0.660867
A930001K06Rik*	74000492	0.658554
Gpr161	74003403	0.656807

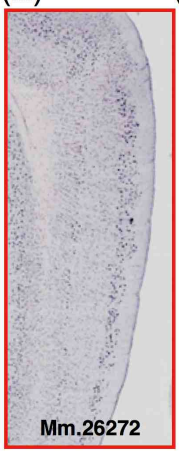


Figure 3.8: *Using NeuroBlast to find all layer-enriched genes* NeuroBlast results for (i) the input gene Hcn1 (raw ISH image) are shown (ii). The correlation score reflects the degree of similarity between genes within the 'entorhinal area'. (iii-iv) Raw ISH images for two of the genes shown in (ii) emphasise that genes with similarly high correlations can show different expression patterns

3.3.3 *Using Allen Brain Atlas 2D image data to develop a resource for extracting complete enrichment lists and further patterns on expression*

The second approach made use of the raw 2D image data rather than ABA tools.

3.3.3.1 *Obtaining images from the Allen Brain Atlas*

Images were downloaded from the Allen Brain Atlas database using the application programming interface (API). First, a list containing the name and image series of all ISH experiments was downloaded from the API. The image series is a unique identifier as ISH was performed and imaged for some genes multiple times, sometimes using multiple probes. Second, for each image series, a list of codes for each image in the series was identified using xml parsing from the file located at: "http://www.brain-map.org/aba/api/imageseries/[iseries].xml". Third, images for download were selected based on their medio-lateral location. The downloaded file also provides information on the approximate medio-lateral location (based on its registration to the reference atlas images: see Fig. 3.2(i)) and path to each image in the series. Using Allen Brain Atlas coordinates, which differ slightly from those of Paxinos and Franklin (2001), MEC is located between approximately 3.125 and 3.5mm laterally. Since the ABA sagittal reference atlas begins at 3.925mm laterally, images between 0 and 1400 µm (refers

to distance from most lateral point) were selected for download for each image series. This range was large to ensure that poorly registered images (i.e. aligned to reference atlas sections that were in fact far too lateral or far too medial) were successfully downloaded. Since most experiments were sampled at 200 µm, 6-7 sections were typically downloaded for each gene, although this varied between genes, as some sections were missing due to sectioning errors or poor tissue quality.

To download each image, the path for each image satisfying the approximate location requirement was extracted from the image series file and included in the following image path: " http://www.brain-map.org/aba/api/image?zoom= [zoom]& top=[top]& left=[left]& width=[width]& height=[height]&mime=[mime]&path=[path].'". The 'zoom' parameter refers to the resolution. This satisfied two constraints: sufficient resolution to identify the density of expression within individual layers, and small enough to keep image processing times to a reasonable limit. The parameters 'top','left', 'width', and 'height' are the coordinates of a bounding box used to crop the original image. These were set to include the entire brain image. The 'path' refers to the specific image path for each image in each image series. Two files were downloaded for each image: an ISH image file and a corresponding expression image file. Images were approximately 500kb each. Approximately 120,000 were downloaded in total and they were stored on a cluster provided by the Edinburgh Compute and Data Facility (ECDF) for processing.

3.3.3.2 *Image pre-processing for subsequent analysis steps*

Images downloaded from the ABA did not all have the same dimensions, so first, images were downsized by a factor of 1.5 and pasted onto the centre of a new image of 1200 (width) x 900 pixels (height) using the Python Image Library. Images were processed using a background subtraction algorithm provided by the image processing software, ImageJ (Schindelin et al., 2012)(Fig. 3.9 (ii)), with radius set to 1 pixel, which is approximately the size of a cell at the chosen resolution. Images were then thresholded such that all visible objects in the image, including anatomical features and cells with very low staining, were retained (Fig. 3.9(iii)). The aim of this step was to minimise gene expression-specific information in the images whilst retaining anatomical detail. Finally, a Gaussian filter was applied to the images to smooth them prior to processing because this aids in feature extraction.

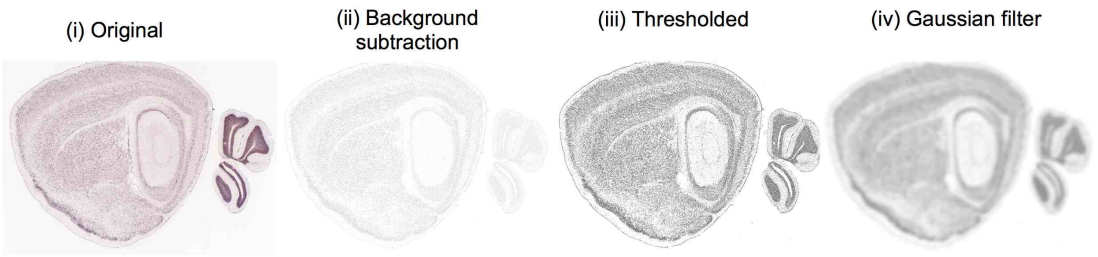


Figure 3.9: *Image pre-processing in preparation for subsequent processing* (i) Original ISH images showing the blue/purple precipitate formed as a result of the alkaline phosphatase - BCIP/NBT stage are converted to grayscale and (ii) the background is subtracted using a rolling ball algorithm with pixel radius set to 1 pixel (approximately the size of a single cell at the resolution used). (iii) Images are thresholded so that all non-background (cells and anatomical features) pixels are converted to black and all background pixels to white. This minimises differences in the expression levels of individual cells, since the intensity of all cells is effectively made either zero or 255. However, cells that previously had stronger staining are likely to appear larger after thresholding. For some processing steps, images are subsequently Gaussian blurred (iv) for optimal feature extraction.

3.3.3.3 *Image segmentation to remove the cerebellar region*

I decided to use an automatic segmentation function to remove the cerebellar region from images prior to registration. First, the cerebellar region is not relevant to this project, and second, a registration algorithm might try and match features in the cerebellum as it contains many prominent features, rather than focusing on the MEC and other regions of interest. If the cerebellum and main brain were out of alignment in images, possibly due to cutting misalignment, then this could compromise the registration of other regions.

Process of image segmentation Prior to segmentation, the images were pre-processed as described. I used edge detection to emphasise the border between the forebrain and the cerebellum. Images were smoothed then an edge detection algorithm was applied (FeatureJ Edge detection (Meijering, 2013) (Fig. 3.10(A)). This image was thresholded to provide a clear brain region and a cerebellar region, sometimes with internal gaps caused by very low expressing brain regions. I used an ImageJ algorithm to identify objects (defined by a complete perimeter) with such internal gaps and to fill in the holes, which typically gave a complete brain region segregated from a complete cerebellar region (Fig. 3.10(A)(iii)). These regions could then be detected as separate objects, using a particle analysis tool, and only the largest object, corresponding to the brain was subsequently included in the mask. This mask was applied to the original image by creating a new image in which only white pixels in the mask were included from the original image.

Automated analysis of image segmentation success Segmentation failed for images where the ISH labelling was very poor (because of edge detection failures), where the cerebellar region and brain region overlapped (due to mounting errors), and where erroneous staining prevented typical boundary detection. It was not feasible to examine all images and check those in which segmentation had failed, so I developed a method for automatically detecting successful segmentation.

To obtain a measure of segmentation success for each image within an image series, I used a binary support vector machine (SVM) classifier with a linear kernel to classify image masks (see Fig 3.10(A)(iv)) based on success. To classify images, it is first necessary to extract features of the image that represent the patterns found in them. I used the VLFeat toolbox in Matlab (Vedaldi and Fulkerson, 2008) and custom-written code (Vedaldi and Zisserman, 2013) to extract scale invariant feature transform (SIFT) features from the images. The extraction of SIFT features involves transforming an image into a large collection of feature vectors, each of which is invariant to image translation, scaling, and rotation, partially invariant to illumination changes and robust to local geometric distortion. SIFT features are invariant to scale and rotations, so are very useful for non-aligned data. The toolbox extracts these features at 4 different scales to provide a spatial histogram of features that contains information about the positioning of features in space. For each image a feature histogram containing 4000 values was generated for input to the classifier. This process is shown in Fig 3.10(B).

I trained the SVM classifier on feature vectors from 800 positive images and 250 negative images and kept a further 800 positive and 250 negative images for validation and tuning of the regularisation parameter (the model was not trained on these images). The classifier was able to separate positive and negative validation images with over 98% accuracy with a tuned regularisation parameter. I therefore used the SVM model with the same parameters to obtain a score for all remaining 120,000 images that estimated their likelihood of being successful or not. The majority of images were assigned positive scores (Fig 3.10(C)).

Although the SVM accuracy was very high, I wanted to flag all images that received a score close to the decision boundary of zero as being potentially poorly segmented (see example in Fig 3.10(D)). I included any image with a score below 1.

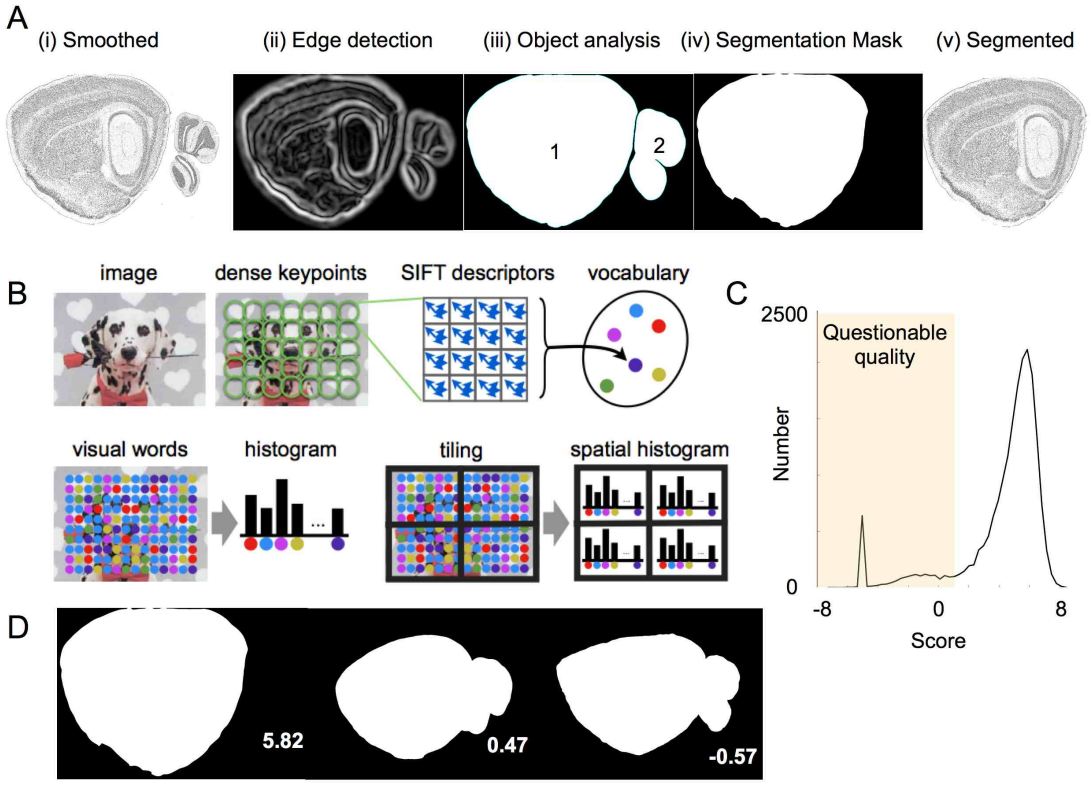


Figure 3.10: *Procedure for image segmentation* (A) Procedure for segmenting images. (i) Smoothed images are processed using an edge detection algorithm (ii) set to detect strong edges. This image is thresholded and regions surrounded by edges are filled in using ImageJ 'Fill Holes'. A watershed tool is applied to segregate regions. (iii) Object detection algorithms distinguish the main region of the brain from the cerebellum to create (iv) a segmentation mask. (v) This can be applied to the original image using ImageJ 'Image Calculator'. (B) Schematic showing the procedure for extracting Pyramid Histogram of Visual Words (PHOW) features from an image (adapted from Bosch et al., 2006). (C) Histogram showing the distribution of segmentation scores assigned by the SVM. All images falling in the shaded area (< 1) were flagged as potentially erroneous. (D) Example images of good and poor masks in which the cerebellar region has not been effectively segmented.

3.3.3.4 *Development of a reference image for each atlas sagittal section containing MEC*

To enable me to extract information from 2D ISH images with precision, I set out to register all downloaded images to reference images corresponding to sagittal sections. First I selected five planes throughout the medio-lateral extent that I felt the MEC and its borders could be captured by and labelled these ML1:5 starting from the most lateral. These correspond to the images used when defining partitions for the AllenMiner (see Fig. 3.6). ML1 contains a very lateral region of MEC, and possibly some regions of lateral entorhinal cortex (LEC). This section was used for reference purposes but not for data extraction. ML2-4 show MEC and all three of the major cell body layers are present. In ML5, the more ventral rostral MEC regions have been replaced by pre- and para-subiculum. Again, this section was used as a reference, but at a later date could be used to extract genes that are selective for one region of MEC, pre- or para-subiculum but not the others.

To form each reference image, one possibility was to register each image to the Allen reference atlas images or reference Nissl images corresponding to the chosen planes. However, these images suffer limitations, first in that they are not of the same media as the ISH images,

because they have been processed and stained differently, and second because they are just from a single mouse brain, and therefore do not capture the variability in the population. As such, I made new reference images using existing ISH images. By registering multiple ISH images to each other, then finding a median of the result, both the limitations of different media and single mouse were removed.

Images were selected for group registration based on (1) relatively uniform expression in the MEC, (2) good tissue quality, and (3) medium ISH staining intensity. Approximately 15-20 images were chosen for each of the 5 reference images (See Fig. 3.11(A)) and were processed to provide optimal input to the registration software. This involved background subtraction and thresholding, followed by rigid alignment. An ImageJ plugin was used to perform the rigid alignment (Schindelin et al., 2012). This allows the user to mark 2 corresponding points on images to be aligned then aligns the images so that these points are moved into the same location.

Images were group registered using a Matlab library, the Medical Image Registration Toolbox (Myronenko and Song, 2010). This toolbox provides several tools, one of which allows a number of images to be registered by two-dimensional non-linear deformation to one another. The aim is to find the group match with the best similarity measure. There were a number of choices to make in using this tool: (1) the number of images to use as input; the relevant issues being accuracy of registration offset against time and memory taken to perform the registration, (2) the similarity measure to use, (3) the alpha value, (4) following registration, use of the mean or median of the group-registered images and (5) the pre-processing of the images.

1. Image number: I found that 15-20 images provided a mean/median image that captured the variance across Allen Brain Atlas images well, without requiring excessive memory (~ 5GB RAM) or time (~ 20 hours) on a cluster.

2. Image pre-processing: Images were pre-processed and segmented as described in the previous sections. A Gaussian filter with a window size of 13 and a standard deviation of 3 was then applied iteratively three times to enhance image structures at the relevant spatial scale.

3. Alpha value: This sets a limit on the scale of the deformation. I reduced this from the default value to prevent large deformations.

4. Use of the mean or median: The mean and median images showed high similarity but the median captured slightly more anatomical detail in the cortex.

5. Similarity measure: I chose mutual information since this is relatively resistant to difference in contrast.

The output of group registration is a series of transforms that correspond to each image. Applying these transforms, followed by calculating the median of the transformed images provides a reference image. This process can be applied to all relevant reference groups (See Fig. 3.11(B)(a)).

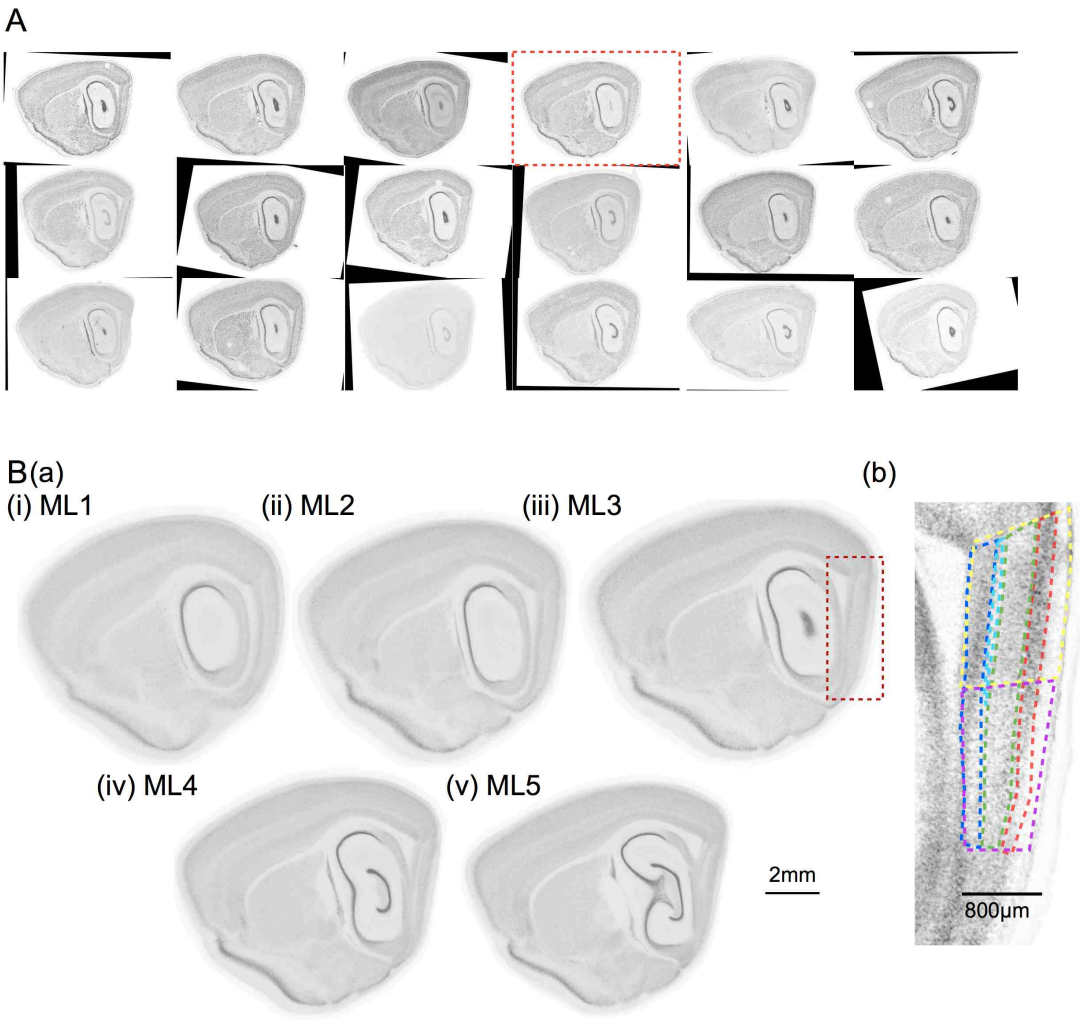


Figure 3.11: *Group registration of images to form a reference image* (A) Images used to form the ML3 reference image. Images have been rigidly aligned such that they are oriented in the same direction. Black regions show the pre-alignment image location, therefore showing the translations and rotations necessary to approximately rigidly align all images to the highlighted image. (B)(a) The median images formed from all group-registered images for each reference group (i-v). The red square indicates the area shown at higher magnification in (b). (b) Manually determined layer boundaries are shown with dotted lines on the ML3 median image. LII: red, LIII: green, LV/VI: dark blue, LIV: light blue, Dorsal: yellow, Ventral: purple.

3.3.3.5 *Delineation of layers based on reference image*

After group registering five sets of images to give 5 reference images, I used reference atlas images as well as the patterns of staining in the averaged reference images to divide the MEC region into non-overlapping regions representing layers II, III,IV and V/VI (Fig.3.11(B)(b)). I also divided the MEC into two regions along the dorso-ventral axis for the purpose of coarsely estimating differences in gene expression. Divisions were achieved by manually outlining the regions using Bezier lines in ImageJ, then creating image masks of white on black pixels to mark out the coordinates of interest for each region. These image masks can be used by Matlab and Python scripts to select particular regions of image arrays.

3.3.3.6 *Classifying images based on their medio-lateral location*

Once reference images had been constructed, the images downloaded from the Allen Brain Atlas could be assigned either to one of the five reference image groups or to a sixth group for images not containing MEC. To minimise the manual effort of classifying each of the 120000 images, I used a Support Vector Machine (SVM) library for Matlab (libsvm) with a linear kernel to make a binary classification on whether an image corresponded to the reference image containing the most complete MEC region, ML3 (Fig. 3.12). The SVM provided a score for each image across all image series reflecting the likelihood that the image was approximately in the same medio-lateral plane as ML3. This enabled me to subsequently compare scores for all images that had been downloaded for a given image series and choose the image with the highest score as the ML3 image. Images from each image series that closely anatomically matched the other reference images could then be identified based on their estimated medio-lateral location (according to the API database file mentioned previously), since images within image series were always separated by 100, 200 or 400 μm.

The procedure for the SVM followed several stages. First, before inputting data to the SVM, all images were pre-processed as described previously. A Gaussian filter was then applied, with a window size of 15 and a standard deviation of 3 to enhance image structures at the relevant spatial scale. PHOW features, which are SIFT feature descriptors extracted at different spatial scales (Fig. 3.10(B)), were extracted from each image to provide input to the SVM. The first stage of using an SVM is to train the classifier (Fig. 3.12(i)). I manually classified images for approx. 500 genes into ML groups to provide training data. A further 483 genes were used for the validation set. This gave 501 images in the positive training data set, and ~ 1573 in the negative training set, with 1600 in the negative validation set. The negative training set was larger than the positive training set because most images are not in the plane of ML3, but it was not four times the size of the positive set because some image series did not have a complete set of images for the 5 chosen reference images. They may have been removed for the ABA due to poor tissue quality or ISH staining. The validation set was used to optimise the parameters for the SVM, such as the regularisation, C, parameter (Fig. 3.12(ii)).

After training, all the remaining images were run through the SVM and assigned a score (Fig. 3.12(iii)). The highest scoring image from each image series was then assigned to an ML3 folder for manual inspection. Any images that did not belong in ML3 were removed. Images that were distantly located from ML3 images were placed in a No-ML folder. I manually checked this folder and any images that appeared to match a reference image were moved to the appropriate folder. Records were kept of all movements and this process of checking continued until I was satisfied that images had been assigned to a reference with ~ 95% accuracy. It was important for this stage to be accurate so that when images were registered to a reference images, the features could be accurately aligned. It is possible that some image series were lost at this stage due to images being misclassified as not belonging to an ML group when they in fact did.

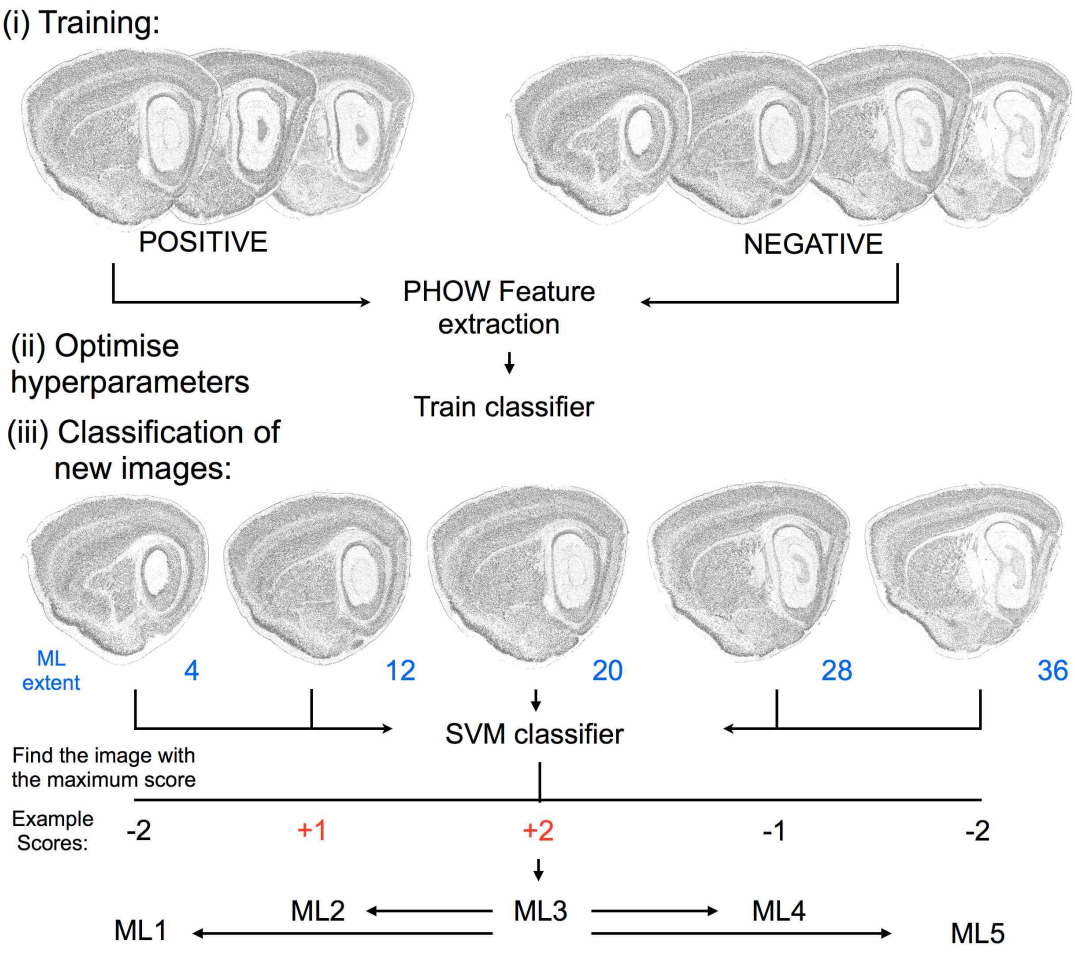


Figure 3.12: *Classifying images based on their sagittal location* Schematic showing the procedure for using an SVM to classify images into medio-lateral group for registration. (i) Training: The classifier was trained on PHOW features from images that had been manually assigned as ML3 (positive) or ML1,2,4,5, none (negative). (ii) Optimisation step to tune hyperparameters for optimal separation of positive from negative images. (iii) The classification of new images step. Images at different medio-lateral extents (see blue number assigned to image) in each image series were input to the classifier and assigned a score (positive scores as shown in red). The highest score could be used to classify them as ML3. The other images were then assigned to a reference group depending on their medio-lateral extent (blue). Images in a series are typically separated by 200 µm, which is 8 x 25 µm, the thickness of each section.

3.3.3.7 *Registration of all images to a reference image*

Once downloaded images had been assigned to appropriate reference ML groups, they could be 1-to-1 registered to their respective reference image. Here, I also made use of the MIRT toolkit (Myronenko and Song, 2010), but this time rather than using group registration, all images were registered to the same median image (Fig. 3.13). 1-to-1 registration may not be as accurate as group registration but group registration would be unfeasible, in terms of memory and time required, for the number of images involved (~ 20,000).

Procedure for registering images For 1-to-1 registration thresholded images were again used to minimise gene-specific detail. A Gaussian filter was applied to the images to encourage

registration of larger image features rather than individual cells. The same filter was applied to the reference images. Images were registered using mainly default settings. The aim of registration is to minimise the difference between a reference image and the ISH image. Mutual information was used as a similarity measure because of its invariance to differences in contrast; ISH images can show large variability in contrast. The output of the algorithm was a transform describing the deformation of all points. This transform, calculated using the thresholded images, could then be applied to the original ISH images (for visual assessment of registration success), as well as to the expression images (for extracting pixel intensity). All images underwent the same procedure (Fig. 3.14).

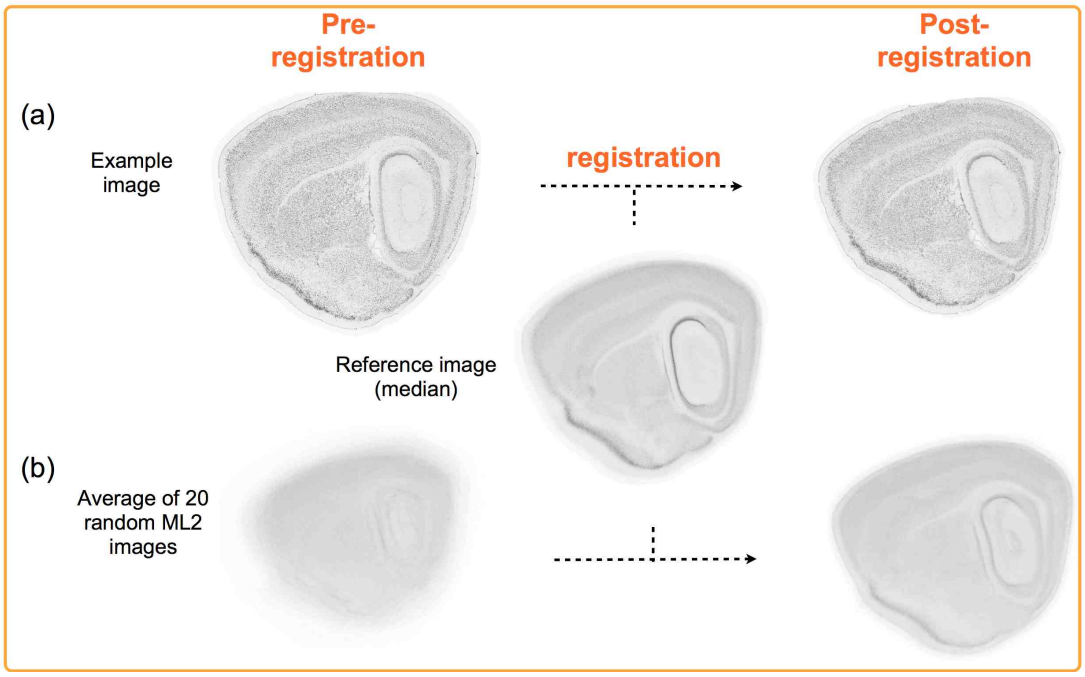


Figure 3.13: *One-to-one registration of each image to its respective reference image* (a) The procedure for 1-to-1 registration is shown for an example gene. The image was registered to its appropriate reference image and the transformation applied to the original (non-thresholded) image (see also Fig.3.14) (b) The effect of registration is shown for a group of 20 random genes. Before registration, when 20 images from random genes are overlaid and a mean calculated, the resulting image lacks structure. After registration, the mean image of 20 shows similarity to the reference.

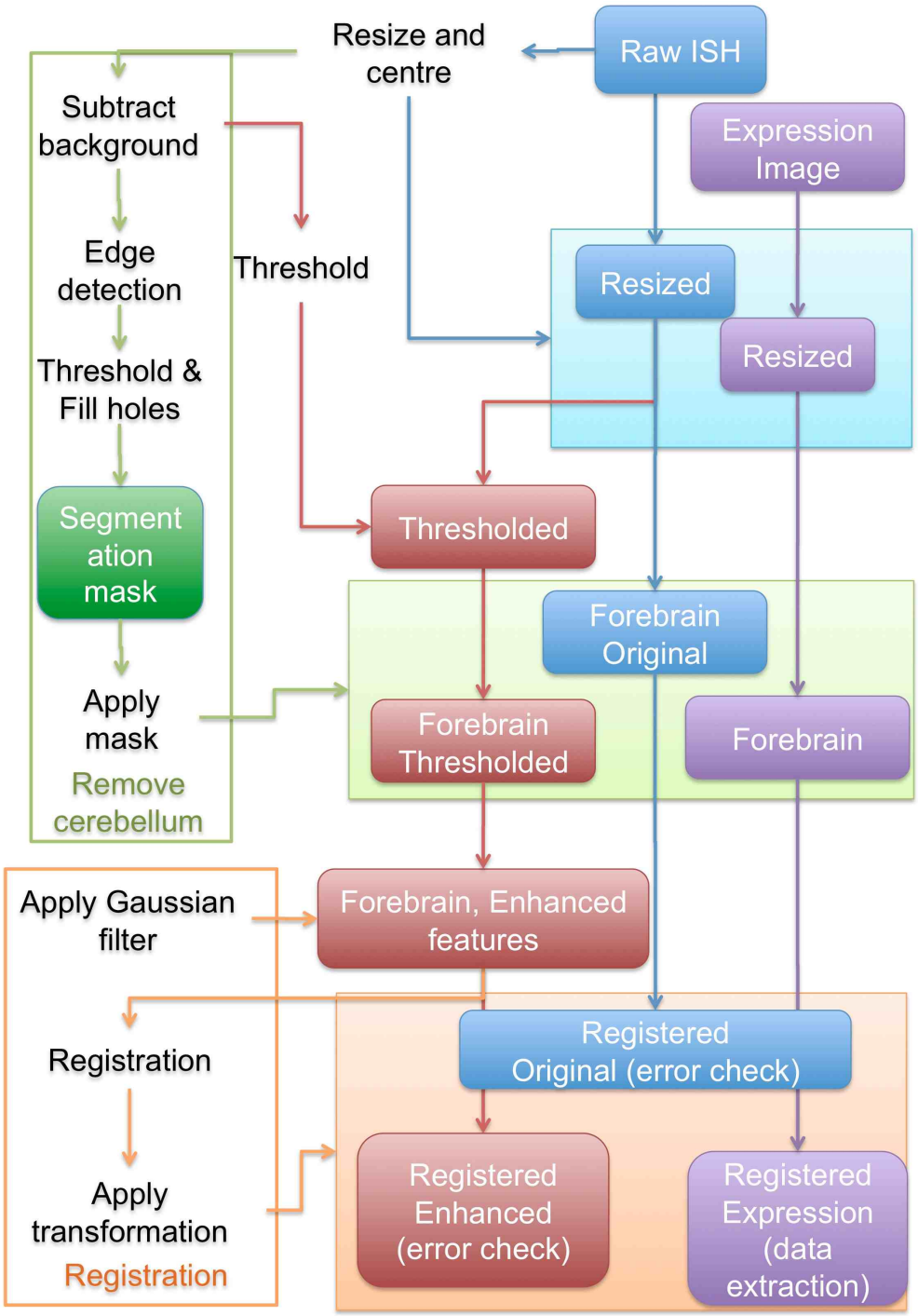


Figure 3.14: *Image processing steps* The process undergone by each type of image from the original, through image segmentation (green) to registered (orange) is shown. Registration is performed on the thresholded, enhanced feature, version of the ISH image, and the transformation is applied to the original ISH image, for validation, and to the expression mask for data extraction.

Procedure for quantifying the accuracy of registration This type of data analysis relies heavily on accurate registration. It was possible to assess the accuracy of registration using several

measures that could be automatically collected from all images: the mutual information score of the registration, a cross-correlation score on the final image, and a classification score from a classifier trained on poorly registered images.

First, I examined the mutual information score of the final deformation for each image during the registration process, which reflects its similarity to the reference image. The distribution of scores before and after registration in shown in Fig. 3.15(A)(a). To determine how well these scores represent correct alignment, I chose a random sample of 100 images and manually rated their registration accuracy, then plotted their scores against the final mutual information result (see Fig. 3.15)(A)(b). This allowed me to determine a threshold so that I could flag images that were potentially poorly registered.

The main limitation of the mutual information score is that is reflects registration accuracy across the whole image and therefore could have over-or underestimated precise registration in the MEC region. As such, the second test I performed was to cross-correlate the MEC-containing region of a registered image with a larger region of the respective reference image. I used a Matlab function (normxcorr2) to find the correlation between the two images at all the different positions the images can be overlaid in (Fig. 3.15(B)(a)). This cross-correlation function provides both a normalised maximum fit score between the two images, as well as the location at which it occurs, thereby enabling me to estimate the offset between the posterior MEC border within the registered image and within the reference image. As for mutual information, I identified meaningful scores by comparing them with visual assessment for the same sample of images. A small proportion of the sample showed non-zero cross-correlation scores (Fig. 3.15(B)(b)) but the majority (see heatmap) were given a high manual rating of 5 and cross-correlation offset of near-zero, thus highlighting the usefulness of this score. However, a number of images given a manual rating of 0-4 were also given a cross-correlation score of zero.

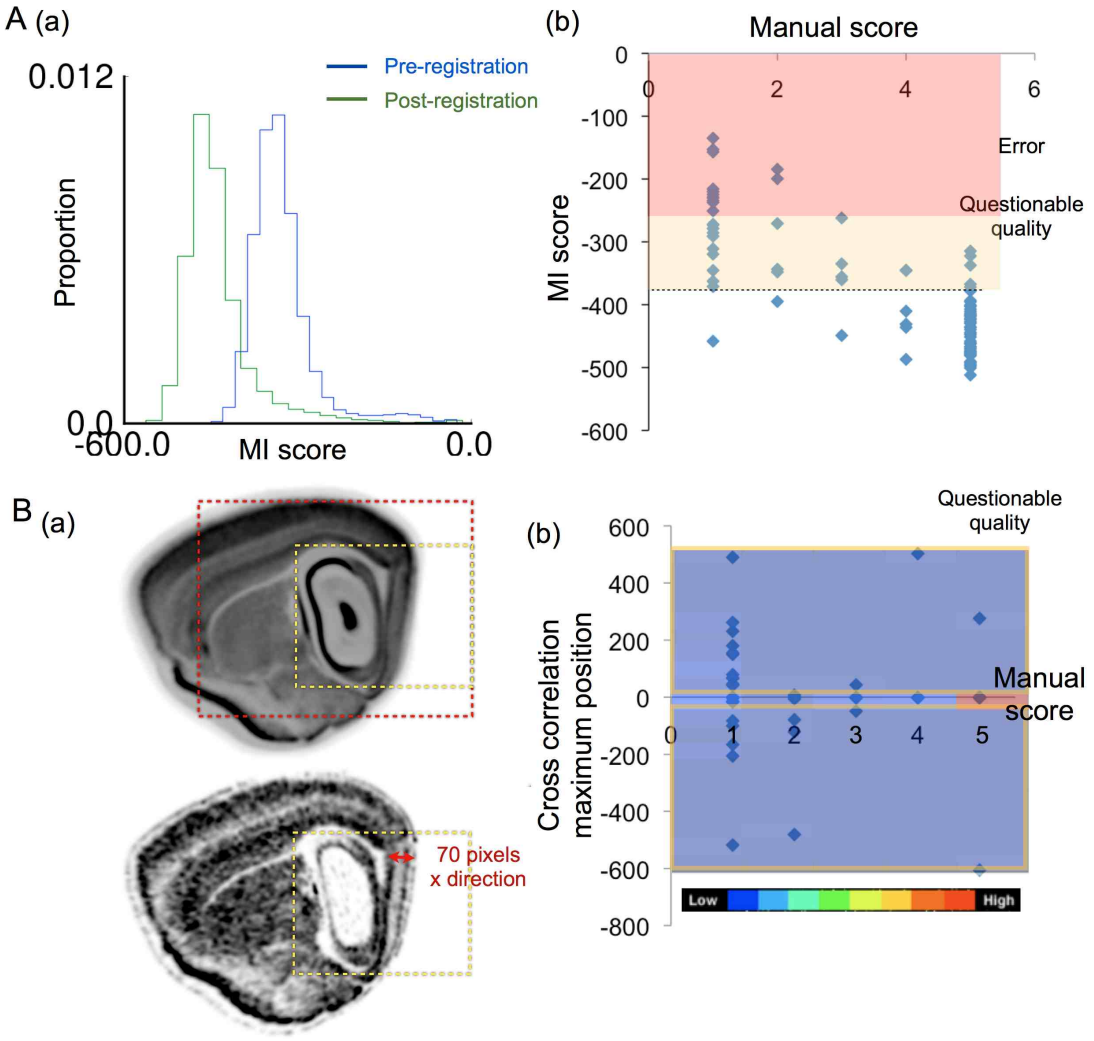


Figure 3.15: *Quantifying the accuracy of registration* (A) Mutual information measures (a) Graph showing the distribution of mutual information scores (output from the MIRT package) before (blue) and after (green) registration. (b) The final mutual information score is plotted as a function of manually judged registration score. Manual thresholds have been applied to judge erroneous (red) and questionable quality (yellow) images. (B) Image cross-correlation measures (a) The regions used as the source (red) and template (yellow) are shown for the median image (upper), and an example gene (lower). In cross-correlation the yellow region is moved pixel by pixel around the source region and the similarity between the two images is recorded for each location. (b) Scatter plot showing position (in the x direction) of the highest cross-correlation score plotted as a function of manually judged score, overlaid on a heat map. Images with good registration should have zero offset. Manual thresholds have been applied to judge images of questionable quality (yellow outline)

3.3.3.8 *Final error classification method*

I used one further measure to identify potentially erroneous images aside from checking segmentation, mutual information score and cross-correlation. The aim for this measure was that it should be sensitive to image flaws including holes in the tissue created by bubbles, and aberrant detection of the pial surface as an RNA-expressing cell, which artificially increases the mean pixel intensity of the image. Examples of flaws in ISH images at the ISH stage and at

the expression mask stage are shown in Fig. 3.1(C). These types of errors must be detectable if data is to be quantified directly from these images.

I decided to use image features to capture erroneous elements in the images that could subsequently be detected using an SVM. PHOW features were extracted from pre-defined regions of the entire image, one including MEC and its surrounding region, and the other including just the MEC (Fig. 3.16(A)(a)). PHOW features were extracted using the Matlab toolbox VLFeat, which extract scale invariant feature transforms (SIFT features) at different spatial scales to form a spatial histogram for each region.

Using the image features extracted from the images of the immediate region surrounding the MEC, for each reference group I trained a binary SVM classifier with a radial basis function kernel on all the images that I had visually assessed that had significant errors (n=~ 50) against high quality images (n=300). I used cross validation to optimise the regularisation parameter, C, and the hyperparameter of the radial basis function, gamma. I then tested all images with the classifier, which provided a probability estimate that an image had an error. I again compared the probability estimate with visual assessment of a subset of the images used previously to estimate accuracy for the MI score, and flagged images with a score greater than my confidence threshold (Fig. 3.16(B)). The classifier distinguishes large registration errors and pial surface errors from high quality images (Fig. 3.16(C)). However, this method is naive to anatomy and not particularly sensitive to minor misalignment errors as it extracts features from the entire MEC region that are scale and alignment-invariant.

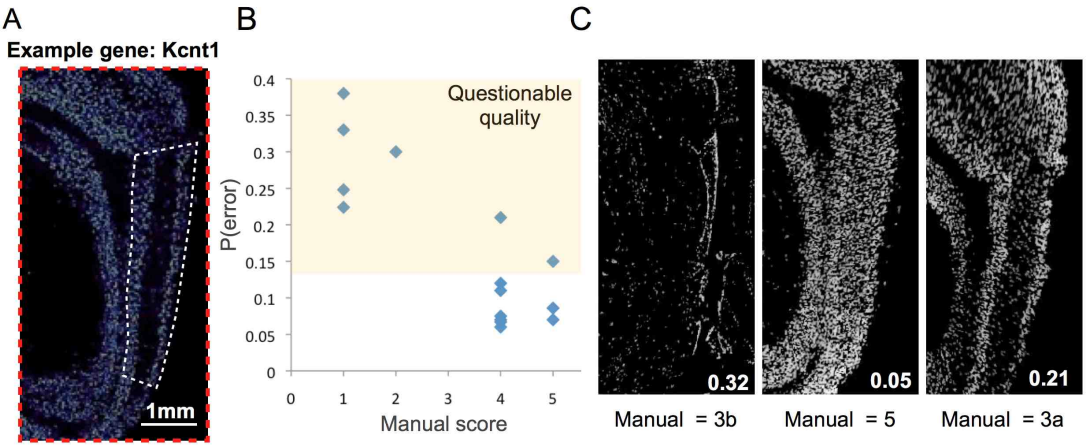


Figure 3.16: *Detecting poor quality images* (A) Expression image overlaid with area indicators to indicate the MEC region (white) and the surrounding region (red). (B) Scatter plot showing the relationship between manual image score and probability that the image has an error. Manual scores were assigned based on this scale: (1): Failed registration - not MEC, (2) MEC shifted, other errors, (3) MEC shifted or other errors, (4) Minor error or shift, (5) No problems (B) Example 'test' images and their SVM probability of error score and respective manual score.

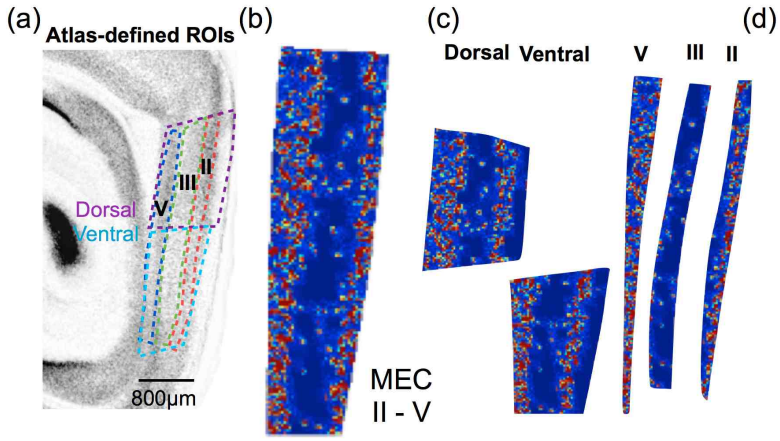
3.3.3.9 *Extraction of pixel intensity data for identification of patterns of enrichment*

I extracted pixel intensity data from expression images (after the registration transformation had been applied), rather than from the raw ISH images because the expression images have already been through stages of processing that control noise, background illumination and contrast invariance across the images (See ABA data production). Further to this, they contain

information about the overall expression level of individual cells that have been detected as expressing the gene of interest and do not contain the structural information present in brightfield images, such as densely fibrous regions, that is not gene-specific.

I extracted data by MEC sub-region. First, regions including layers and dorso-ventral regions were defined in the reference image based on known anatomy (Fig. 3.17(a)). Custom-written Python scripts were used to extract all pixel values at specific coordinates from each region separately (Fig. 3.17(b-c)), and to calculate particular first-order summary statistics, including mean expression, the standard deviation in expression, the mean expression in all non-zero pixels (i.e. to extract data on cells), as well as the median expression and quartiles (Fig. 3.17(d)). The script also calculated the number of pixels exceeding specific thresholds, for example, the mean pixel value + 2 standard deviations, for analysis of high intensity pixels.

This method relies on accurate anatomical delineation and very high quality and precise registration, but it is straightforward to use this information to define regional expression.



(d)

Region	Mean	SD	% Pixels > thresh
Whole	24.4	39.8	
Dorsal	20.7	35.9	49
Ventral	21.3	39.0	51
V	37.8	43.7	51.68
III	7.7	20.9	6.24
II	30.3	45.1	42.08

Figure 3.17: *Extracting pixel intensity data from expression images.* (a) The median reference atlas image for ML3 is shown cropped to the MEC region and is overlaid by the manually defined borders (borders defined using the reference atlas as a guide). Pixel values for (b) the combined MEC cell body layers and (c) for individual regions were extracted from the image. (d) Pixel intensity statistics are shown. The gene is expressed at higher levels in layers II and V but, at a coarse estimate, does not show differential expression in dorsal and ventral MEC.

3.3.3.10 *Classification of images based on laminar patterns in gene expression*

I used features of pixel intensity data to automatically classify images into groups depending on their laminar expression. These features included summary measures of pixel intensity: differences in the mean and standard deviation across layers, and a feature that relates to the skewness of the data within each layer, the proportion of pixels that exceed a threshold. This threshold was calculated individually for each gene using pixel statistics from the whole MEC region. The reason for not using skewness is that I'm particularly interested in the upper tail of the distribution, which provides information about the laminar distribution of the cells that express high levels of a gene. Skewness takes the entire distribution into account.

This method required two decision phases: one to identify the image statistics that best capture the expression pattern of a gene, for example difference in the mean across layers,

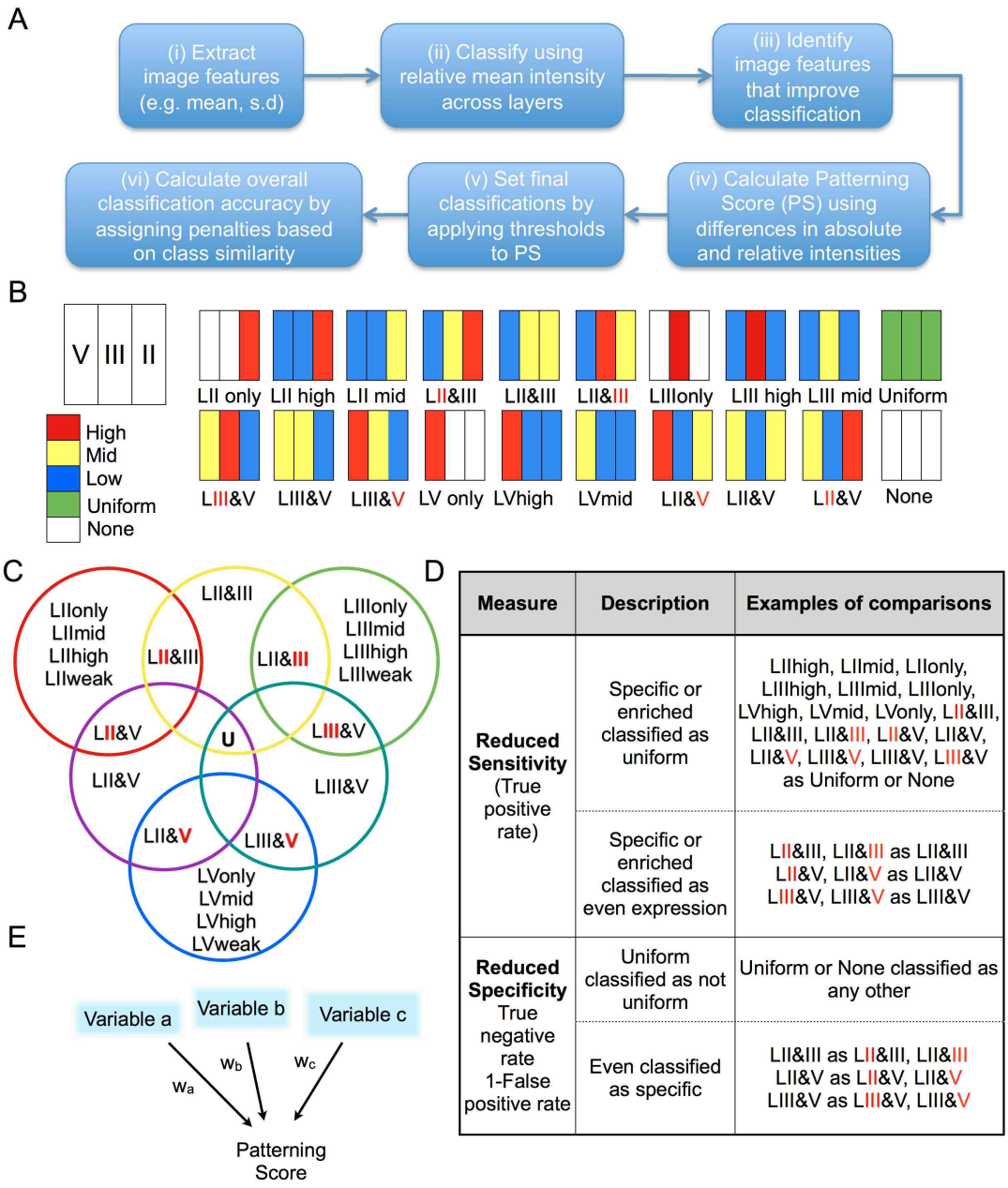
and/or differences in the proportion of pixels exceeding threshold, and the second to assign appropriate thresholds for these features to distinguish classifications (see Fig.3.18(A)). To aid in choosing features and thresholds, I used genes that I had previously identified using NeuroBlast and subsequently validated as training data, and extracted data from their registered images. All genes with a rating greater than 1 in any category were reclassified into unique categories based on precisely which layers they were present in (see Fig.3.18(B)).

In order to find potentially useful features or combinations of features, I extracted pixel data from the training images and examined how well the data from different 'ground truth' categories could be segregated using the different features. I performed a principal component analysis using R, and assessed which features showed the highest loadings onto the principal components and which were redundant. The results are outlined in the Results section.

Once I had identified features, I also used the training data to set appropriate thresholds to segregate the data into the laminar categories (Fig. 3.18(B)). To aid in setting thresholds for each decision boundary (Fig. 3.18(C)), I plotted confusion matrices to reveal the proportion of ISH images whose manual classification matched the predicted classification. I manually adjusted these thresholds until I achieved good performance across all categories. Genes with zero expression were identified using the mean pixel intensity of the MEC region.

In classification, good performance is a somewhat subjective measure and depends on the question being asked. The advantage of using manual classification over a supervised classifier is that it is completely transparent, and classification can be tuned based on the degree of sensitivity (true matches/(true matches + missed matches)) or specificity (true matches/(true matches + false matches) required. Examples of deficits in sensitivity or specificity are shown in Fig. 3.18(D). High sensitivity can lead to Type I errors (false positives), whereas high specificity can lead to type II errors (false negatives). For some purposes it is useful to capture all genes showing any sign of a difference across layers (high sensitivity); in this type of list the number of 'uniform' genes should be relatively low. For other purposes, it is necessary to only identify genes with strong differences, and to minimise erroneous classifications, so this type of list contains a large number of 'uniform' genes.

In order to give the classification scheme flexibility such that genes could be differentially classified depending on question, I assigned each gene a patterning score (PS). I calculated this PS based on key features, weighted according to their importance to classification (Fig. 3.18(E)). Patterning scores were independent of the specific laminar expression. Adjusting the threshold on the PS enables flexible movement of genes both across and within subsets (Figs.3.18(C-D)). For example, if the question is only interested in very high specificity, some of the genes classified as 'LII mid' gene could be moved to 'LII weak' by increasing the necessary PS threshold. The 'weak' categories were brought in specifically to divide up the large 'LII mid' group, which could not all be visually assessed, based on specificity.



Measure	Description	Examples of comparisons
Reduced Sensitivity (True positive rate)	Specific or enriched classified as uniform	LIIhigh, LIImid, LIIonly, LIIIhigh, LIIImid, LIIIonly, LVhigh, LVmid, LVonly, LII&III, LII&III, LII&III, LII&V, LII&V, LII&V, LII&V, LIII&V, LIII&V, LIII&V as Uniform or None
	Specific or enriched classified as even expression	LII&III, LII&III as LII&III LII&V, LII&V as LII&V LIII&V, LIII&V as LIII&V
Reduced Specificity True negative rate 1-False positive rate	Uniform classified as not uniform	Uniform or None classified as any other
	Even classified as specific	LII&III as LII&III, LII&III LII&V as LII&V, LII&V LIII&V as LIII&V, LIII&V

Figure 3.18: *Laminar classification of genes based on pixel intensity values* (A) The procedure for classifying images into the different groups (B) Schematic showing the different types of layer classification. Genes could also be classified as 'weak' if in classified in the 'mid' category but showing a low PS. (C) Schematic showing classification of genes into 23 categories depending on the relative expression of the gene in each of the three layers and the PS score. Red indicates a bias towards one layer when gene expression is present in both. (D) Table showing the types of errors that influence specificity and sensitivity of the classification scheme. (E) Schematic to show calculation of the PS from 3 key features of the images, such as proportion of high intensity pixels in any particular layer. Features are weighted by importance

3.3.3.11 *Classification of images based on gradients in gene expression*

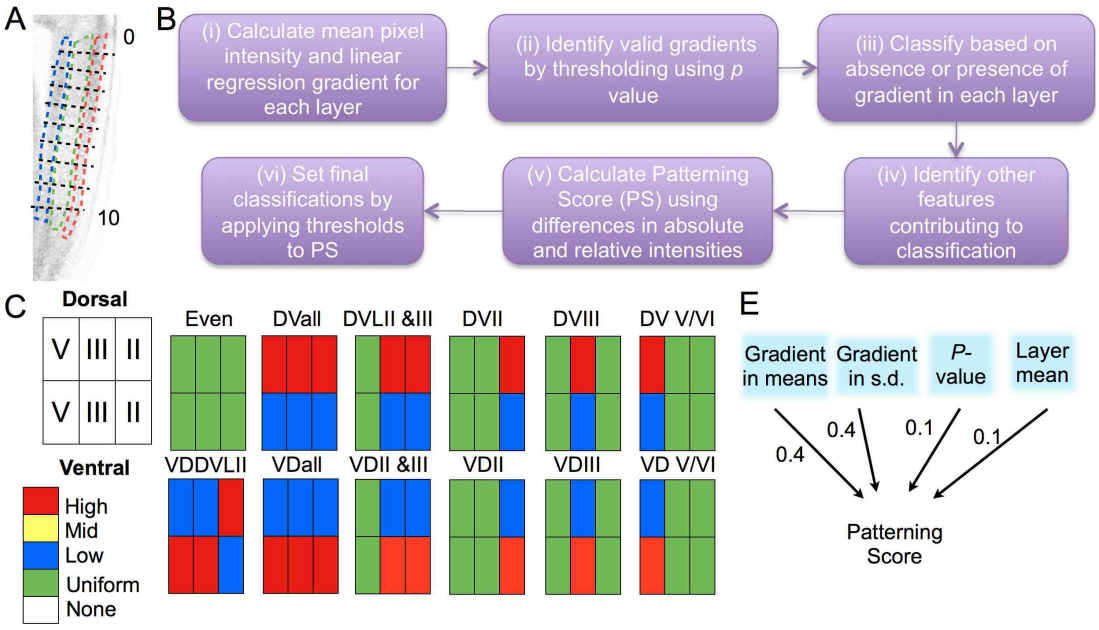
To estimate the presence of gradients I used a different method. I divided the MEC region into 10 subregions along the dorso-ventral axis, and for each layer, calculated the mean pixel

intensity, standard deviation in pixel intensity, and threshold pixels exceeding threshold in each region (Fig. 3.18(A)). These values were used to calculate linear regression gradients for each layer (Fig. 3.18(B)). I used the presence or absence of a near-significance (alpha = 0.3) slope in each layer to classify the images into different groups (Fig. 3.18(C)).

As for layers, I wanted a means of flexibly adjusting classifications to change the sensitivity and specificity of the categories and reduce the chances of the types of error shown in Fig. 3.18(D). This was particularly important for dorso-ventral genes, of which there were many but the degree of gradient varied considerably. I used several features, including the types of gradient, to calculate a PS (Fig. 3.18(E)). Genes that had a sub-threshold PS were classified into secondary and tertiary groups: e.g. $DVII_2$ and $DVII_3$.

3.3.3.12 *Generating finalised gene lists*

The final gene lists were selected such that the most specific groups were relatively small and patterned, whereas less enriched genes were defined as weakly enriched in a particular layer or weakly expressing a gradient. All genes were assigned both a laminar classification and gradient classification. All genes automatically classified as being mid-strongly enriched were manually checked and categorised according to this classification (this excluded LII_3 (weak), uniform and none for layers, and $DVII_3$, Even and None for gradients due to the large number). If the manual check failed, gene images were re-assigned to an appropriate group based on their localisation and degree of patterning. Incorrect classifications were predominantly due to undetected, subtle image errors. Although the aim was to have fully automated classification, the lists of enriched genes were sufficiently small that manual validation was appropriate. Genes classified in the unchecked groups are more likely to be erroneously classified than genes in the checked groups. All calculations were taken using ML3 images, except where there was an error with an ML3 image but not the ML2 image, in which case, the ML2 classification was used.



D

Measure	Description	Examples of comparisons
Reduced Sensitivity (True positive rate)	Gradient classified as uniform	DV II, DV III, DV II & III, DV V,DVall, VD, VDII as Even or None
Reduced Specificity True negative rate 1-False positive rate	Even classified as gradient	Even or None as DVII, DVIII, DVII&III, DV V, DVall, VD, VDLII
	Layer not identified	DVII, DV II&III, DV V as DV all VD II, VD III, VD V as VD all

Figure 3.19: *Gradient classification of genes based on pixel intensity values* (A) Schematic showing the division of MEC layers into 10 regions along the DV axis, overlaid on the reference image. To identify gradients, a linear regression slope was calculated for the mean pixel intensity or proportion of high-intensity pixels for each region. (B) The procedure for classifying images into the different groups (C) Schematic showing the different types of dorso-ventral classification into 6 categories based on whether genes show higher expression in dorsal or ventral regions, and whether there is a dorso-ventral gradient in layer II, III and/or V. (D) Table showing the types of classification errors leading to reduced sensitivity or specificity. (E) Schematic showing the features used to calculate a patterning score.

3.4 RESULTS

The aims of this chapter are to identify the types of pattern that individual and groups of genes are expressed with, and to investigate how gene expression across the whole genome is distributed across MEC. To do this I have used two complementary approaches to extract gene expression patterns from two-dimensional images of *in situ hybridisation* experiments; the first uses tools from the Allen Brain Atlas (ABA) to identify genes that are specifically expressed within MEC, whereas the second extracts lists of enriched genes using a resource developed for the specific purpose of identifying enriched genes within MEC, with confidence, using direct image processing of the 2D ISH expression mask images. This Results section is divided into three subsections: (1) examination of the individual expression patterns of genes, (2) quantification of the accuracy of ABA tools and compilation of an unbiased list of enriched genes identified using these methods, and (3) quantification of the accuracy of the new resource and identification of all genes that are differentially expressed across different layers and regions.

3.4.1 *The expression patterns of individual genes*

3.4.1.1 *Gene expression boundaries correspond with anatomical boundaries*

To identify all genes showing specific expression within MEC, I first investigated whether gene expression corresponds to anatomical delineations such as the layers of MEC. Using Allen Brain Atlas tools including the Fine Structure Annotation, Anatomical Gene Expression Atlas and AllenMiner, I identified a number of genes with exemplary patterns of laminar expression where changes in expression level closely follow anatomical boundaries. Some genes are only expressed in cells of one layer, whereas others are not confined exclusively to one layer, but their most dense expression is consistent with known anatomical borders, as shown in (Fig. 3.20) where the genes are predominantly expressed in layer V/VI or II, respectively.

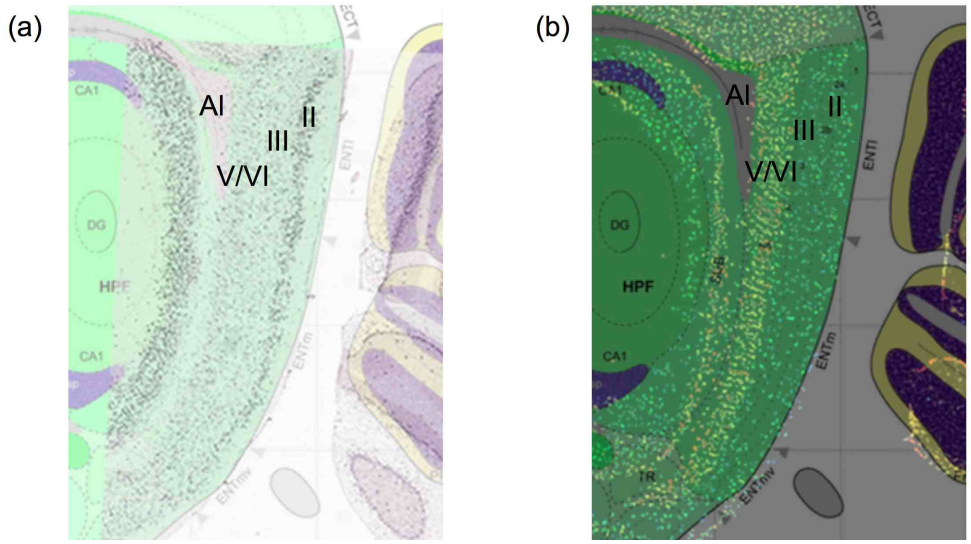


Figure 3.20: *Gene expression follows anatomical layer boundaries* (a) The ISH image for *Hcn1* has been manually overlaid on the Allen Reference Atlas image to show that its expression follows layer II (II). (b) The expression level image for *Kcnmb4* is overlaid to show layer V/VI specificity. Images were aligned using the grey area of the alveus (A). Both images had to be rotated to correspond with the reference.

3.4.1.2 *Many anatomical laminar patterns within MEC are represented by individual genes*

It is possible to identify genes with expression that corresponds to any specific layer or combination of layers. Some genes, such as *Gpc5*, *Cd44* and *Tle4* are expressed uniquely in layers II, III or V/VI, while other genes are expressed at higher levels in these cell body layers than others (Fig. 3.21). A further subset of genes, such as *Cbln4*, *Kcnf1* and *Cd47*, are expressed at higher levels in two of the layers, but not the other layers. Genes with such specific expression are ideal for use as 'model' genes that can be used to find genes with similar patterns using tools such as the ABA NeuroBlast feature. There are also many uniformly expressed genes and genes that show no expression across MEC.

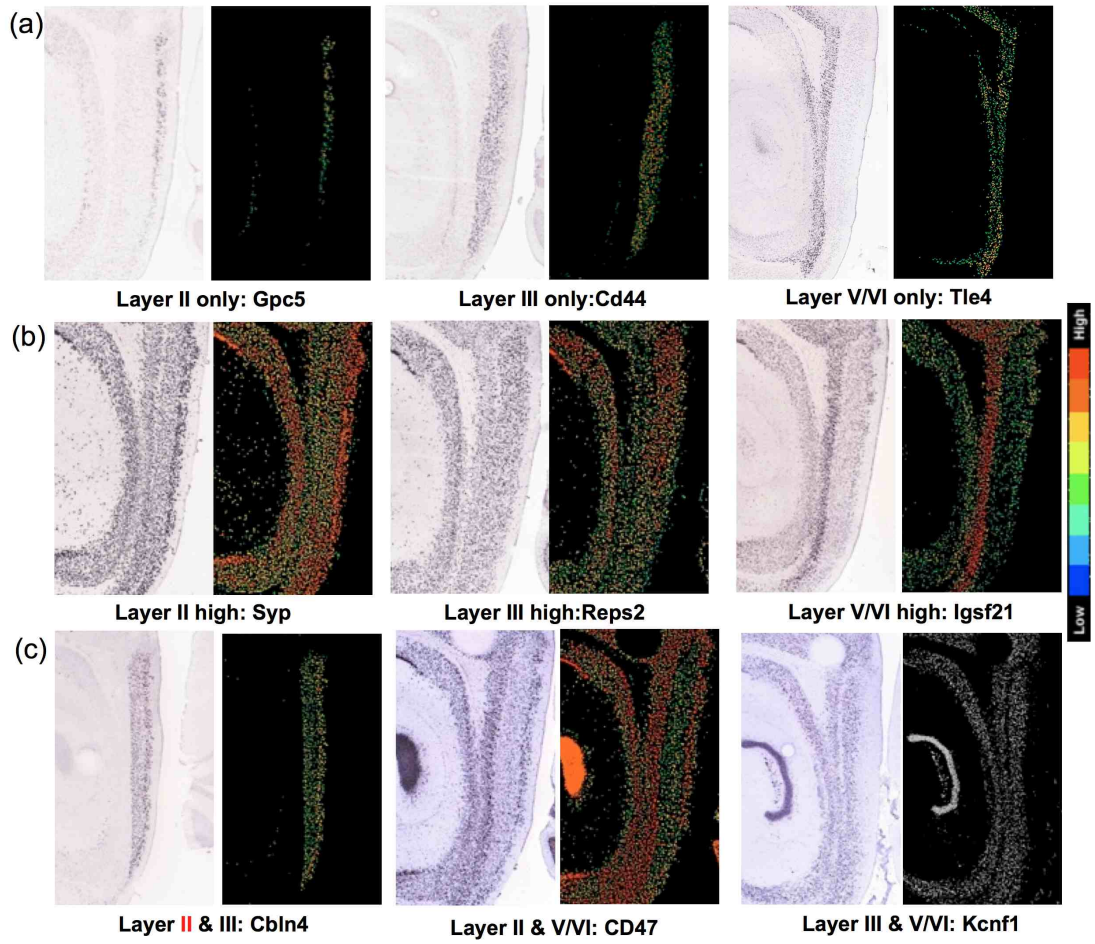


Figure 3.21: *Examples of genes showing enriched expression.* ISH images (left) and expression images (right) are shown for genes with specific expression for each of the categories in which (a) a single layer is enriched in the absence of others, (b) a single layer shows higher expression than other layers, and (c) a combination of layers shows higher expression than the other. Layers marked in red represent a bias towards this layer.

3.4.1.3 *Dorso-ventral organisation of the expression of individual genes*

Genes are not only expressed in patterns that correspond to layers, but also show variation along the axis perpendicular to laminar changes. Genes can be expressed at higher levels in dorsal than ventral regions, across the whole MEC or in individual or combinations of layers (Fig. 3.22). Other genes are expressed at higher levels in ventral regions, but only in layer II or only in layer V/VI, with other layers showing either uniform expression, or the opposing gradient. Thus, gene expression is highly variable across MEC, and shows patterns that are not necessarily consistent with known anatomical or physiological patterns of variation.

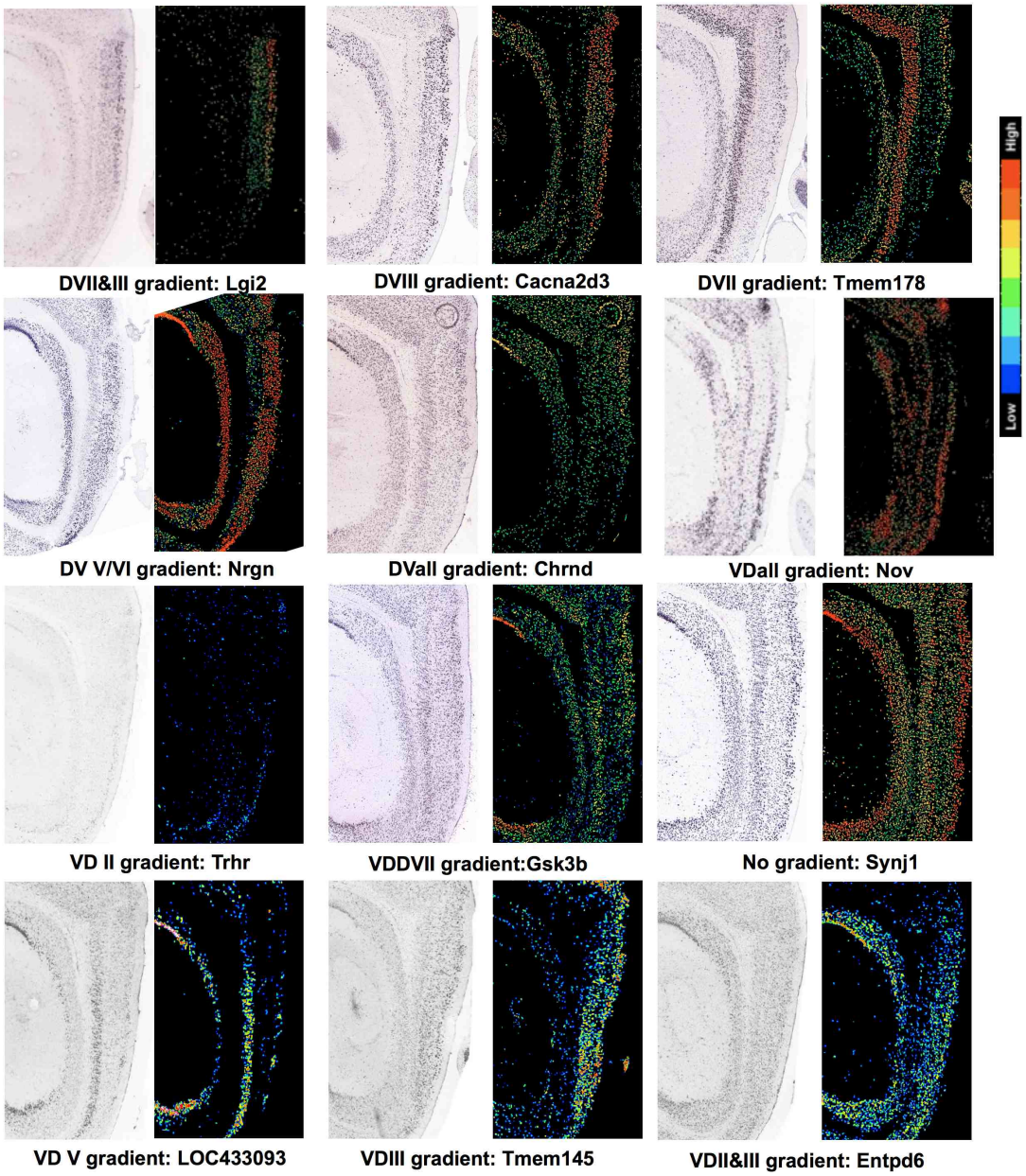


Figure 3.22: *Examples of genes showing a gradient in expression* ISH images (left) and expression masks (right) are shown for genes with either doso-ventral (DV) or ventro-dorsal (VD) gradients in expression that are specific or non-specific to particular layers. The name of each gene is given.

3.4.2 *Using Allen Institute tools to perform genome-wide searches for layer-specific and dorso-ventral patterned genes*

The genes shown above are often not unique in their expression patterns. As outlined in the Methods section, it is possible to identify extensive lists of specifically expressed genes within MEC by first identifying model genes (Fig. 3.21) then using a NeuroBlast tool to identify genes with similar expression. I generated lists of model genes for the groups shown in Table 3.1

and used the NeuroBlast search on all image series (which correspond to ISH experiments) within these lists to identify ~ 5,000 experiments containing images that potentially show specific expression in a particular layer, or with a particular gradient. These images were all subsequently checked for true enrichment.

3.4.2.1 *NeuroBlast predicts enrichment, but with limited accuracy*

For NeuroBlast to be a useful tool, it must both accurately predict enriched genes, and enable calculation of the accuracy with which it enables classification. NeuroBlast predictions correspond with subsequent visual assessment for a subset of genes, but it also predicts a considerable proportion of experiments where RNA is detected uniformly across MEC, is absent in MEC, or is present in a different layer than the model gene (Fig. 3.23(A)). The NeuroBlast tool appears to be most accurate for detecting genes with a dorso-ventral gradient in expression (Fig. 3.23(A): up to 50%), perhaps because of the large volume that this type of change can be detected in.

Laminar results are most accurate for layer II&III model gene searches, where genes are correctly identified as layer II&III enriched up to 30% of the time. Results for layer V genes are considerably poorer, at less than 15%. The specificity of NeuroBlast, which represents its ability to distinguish true enriched genes from uniform genes, is, on average, 15.5% for genes with high correlation scores to model genes (> 0.7). Given that the number of genes with a high correlation is low, I also measured specificity when genes with lower correlations were included in the search. Correlation score weakly predicts accuracy, with specificity falling to 14.3%. I could not directly estimate the sensitivity of the tool, but the low specificity indicates that even though thousands of correlated genes have been identified as candidates, a large number of enriched genes may have been missed.

3.4.2.2 *The laminar expression patterns of enriched genes are dominated by high layer II expression*

Manual validation shows that genes identified using NeuroBlast as being specifically enriched in particular layers of MEC are in fact most likely to be uniformly expressed, or not expressed at all (Fig. 3.23(B)(a)). However, the results reveal a high degree of non-uniformity. Of those genes that are enriched in particular layers, over 800 are enriched in layer II, compared with just 200 in layer V/VI and only 30 in layer III (Fig. 3.23(a)). Fewer genes are expressed in a combination of layers, but not the other, than are enriched in one layer over the others.

To establish whether the non-uniform distribution of enrichment across layers applies to highly specifically expressed genes and genes with combinatorial expression, I reclassified all genes expressed at a medium-strong level into unique groups based on both localisation and specificity (see Fig. 3.23 (B)(b) for unique classifications). While 40 genes are specific to layer II, just 13 genes are specific to layer III and 15 to layer V/VI (Fig. 3.23 (B)(b)). Just 5 genes detected in the NeuroBlast search show expression in layers III and V, but not layer II. This could be a sampling problem as genes with this pattern were not explicitly searched for using NeuroBlast, or it could be the case that this combination is very rare.

3.4.2.3 *Dorso-ventral genes outnumber ventro-dorsally expressed genes*

Laminar differences in gene expression are not the only sources of variability across the MEC. I also observed a considerable proportion of genes showing non-uniform expression across the dorso-ventral axis, predominantly with higher dorsal expression (Fig. 3.23(C)(a)). The majority of these genes (>89%) are expressed with a dorso-ventral gradient in layer II. To establish the distribution of gradients across all layers, I re-classified these genes based on the exact layer in which there is a gradient. The majority of genes that exhibit a gradient (~ 80%) are expressed at higher levels in dorsal than ventral layer II (Fig. 3.23(C)(b)). A large proportion of these genes, ~ 60%, also show a gradient in layer III, with a much smaller proportion (~ 10%) showing dorso-ventral gradients only in layer III or layer V/VI. Of interest, ventro-dorsal gradients are rare, with less than 10% of all genes exhibiting them.

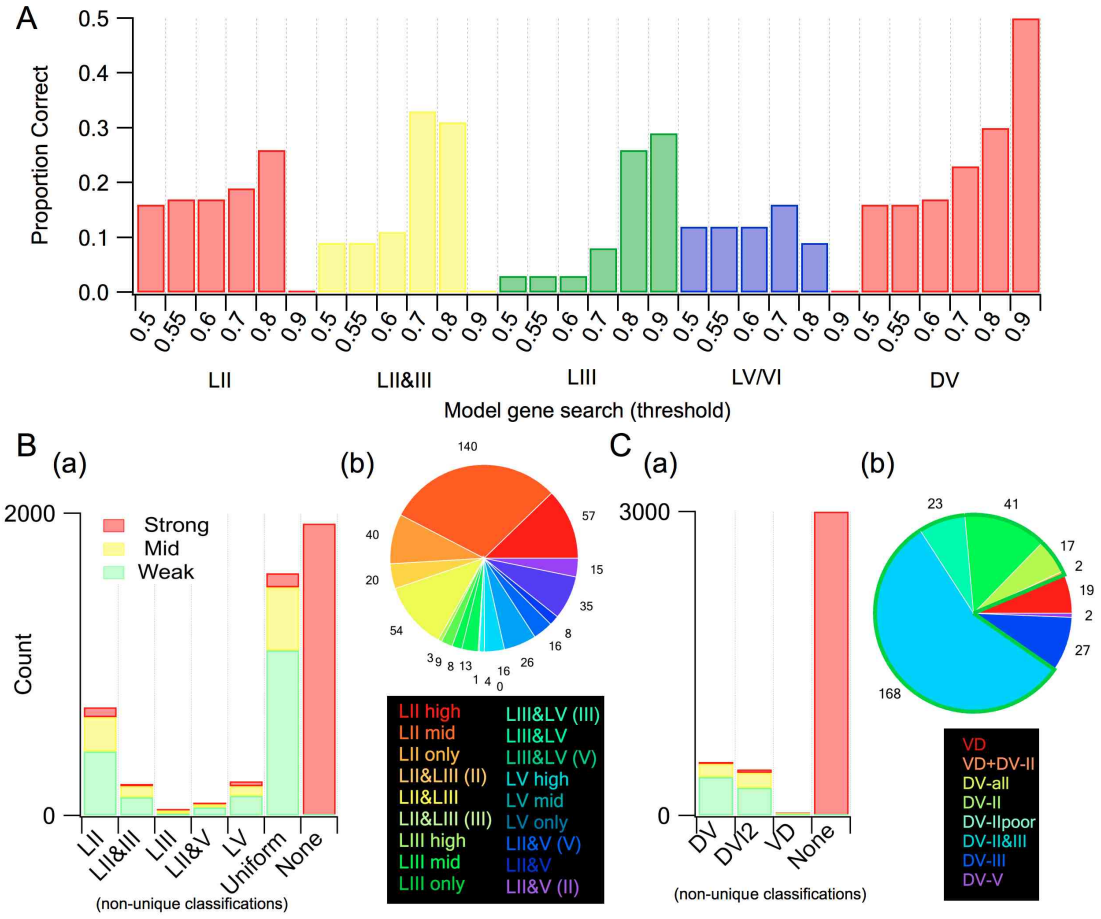


Figure 3.23: *Gene expression distribution in NeuroBlast-detected genes* (A) Bar charts showing the proportion of correct enriched gene hits for each of the NeuroBlast searches on different model lists. The accuracy has been plotted for lists of genes with different threshold correlations to show that results with high correlation are more likely to show correct enrichment. (B)(a) Bar chart showing the number of genes manually classified in the initial identified categories. Colours represent the strength of enrichment. Nb. these classifications are not unique. (b) Pie chart showing the distribution of mid-strongly enriched genes once uniquely classified into layers. (C)(a) As (A)(a) but for dorso-ventral classifications. (b) Pie chart showing the distribution of mid-strongly enriched genes once uniquely classified according to their gradient in expression. The green region outlined shows the proportion of genes showing a dorso-ventral gradient in layer II.

3.4.3 *Identification of gene expression patterns using high-resolution 2D image data*

As the results presented demonstrate, the NeuroBlast tool provides an unbiased means of detecting specifically expressed genes that are differentially expressed within the MEC and of analysing the distribution of expression. However, the major disadvantages of using the NeuroBlast function are the lack of a measure that clearly predicts the validated classification of a group of genes, and the inability to question the similarity of patterns at high resolution. Similarity in the correlation score does not accurately represent how similar the expression of two genes in the MEC truly is. In this section, I'll describe the results of a genome-wide search for MEC-enriched genes that uses registration tools to align all gene expression images, at

10µm resolution, to a common reference. This resource not only provides semi-automated identification of genes that are enriched in a particular layer or region, but also provides a quantitative measure of the degree of patterning of a gene, independent of localisation.

To be able to automatically classify individual genes requires several stages (outlined in more detail in Methods section 3.3.3.9): (1) manual approximation of the position of layers and the MEC borders for pixel intensity extraction, (2) identification of potentially erroneous images, (3) extraction of all pixel intensity values at these MEC locations from registered ISH images, (4) the calculation of image features, (5) the use of previously visually assessed images to identify which features confer the most robust information about patterns of gene enrichment, (6) analysis of the similarity between manual and automated classification to calculate the level of accuracy that can be expected from the use of image features, taking into account possible erroneous features in the images, and (7) final classification of each gene based on laminar expression and the presence of a gradient in expression. I used the reference images and Allen reference atlas to manually define regions corresponding to layer II, III and V/VI.

Analysing the success of 2D image registration As for the ABA tools, this method of extracting pixel intensity information directly from images relies heavily on the accuracy with which alignment can be performed. I both optimised the registration technique to the ISH images and developed several means of identifying poorly registered images, to ensure that I could extract gene expression data with accuracy. These included measuring the similarity between each post-registration image and the reference and using a classifier to detect anomalous images. Even with highly conservative error detection, it is possible to extract high-quality data directly from over 11,000 images from the ABA at the sagittal level corresponding to the ML2 and ML3 reference images (Fig. 3.24(A)). The other images were either flagged or removed from further analysis. Flagged images with predicted enrichment were visually assessed before inclusion in enrichment lists. Many of these images were in fact of high quality, so were reintroduced into the analysis. Images excluded from analysis typically showed very low ISH staining, which limited the number of features that could be detected and therefore registration accuracy.

The accuracy of the registration of the 2D images is high, as demonstrated by plotting average gene expression across all genes satisfying quality criteria before and after registration (Fig. 3.24(B)). Borders within the brain and brain edges are clearly defined, indicating that the alignment of images in the two medio-lateral groups ML2 and ML3 is, on average, very good (see difference in pre- vs post-alignment).

The images in Fig. 3.24(B)(b) also reveal the predominant expression patterns within the MEC, as the post-registration images are entirely a function of gene expression in that no explicit anatomical information has been provided. On average, RNA labelling is higher in MEC layer II than any other layer, and follows a predominantly dorso-ventral gradient in expression there. Layer V/VI can be clearly identified by expression pattern, but the tendency towards increased dorsal expression is not present. Other features, such as the cell body-free layer IV and non-uniformity in the layer V/VI region that could correspond to laminar boundaries can also be identified in average images for these two image planes.

In order to quantify the laminar and dorso-ventral distribution of RNA labelling in MEC, pixel intensities were extracted for pixels in the defined regions and the mean intensity was calculated. The distribution of pixel intensities is very similar across layers II, III and V/VI for pixels with low intensity, but layer II contains the highest proportion of high intensity pixels (Fig. 3.24 (a)). Pixel intensities are also differentially distributed along the dorso-ventral axis of layer II. The most dorsal layer II region contains the highest proportion of all but the lowest pixel intensities (Fig. 3.24 (b)).

A

Stage	Good	Flag	Error	Images remaining
Downlo ad				19358
Segme ntation	17247	200	1911	17247
Registr ation	16152	2300	900	15667
Error check	13710	3700	2300	11786

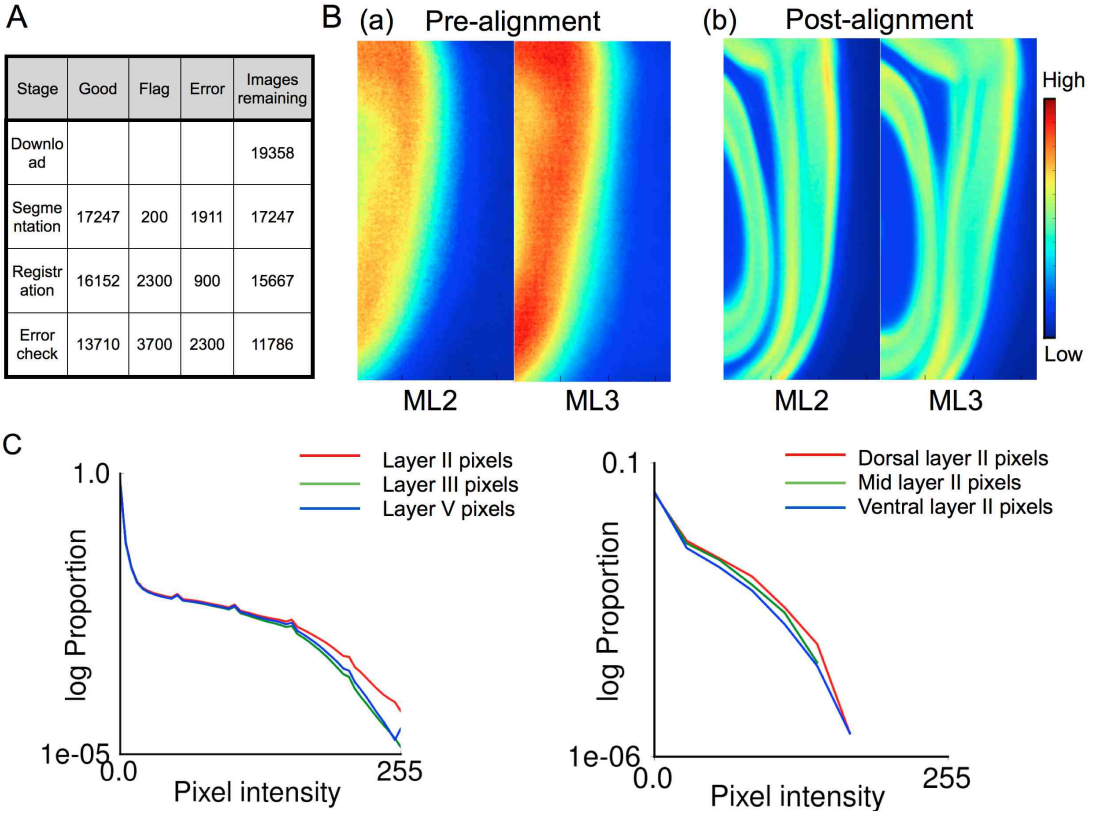


Figure 3.24: *Success of the 2D image registration method* (A) Table showing the number of images that meet quality criteria for the different error measures. 'Flag' images are of questionable quality, but do not necessarily contain errors. (B) Images showing the gene expression levels in ML2 and ML3 images, averaged across all genes (a) pre-registration and (b) post-registration. (C) Histograms show (a) the distribution of pixel intensities across layers (II: red III: green, V/VI: blue) and (b) the distribution of mean pixel intensities across the dorso-ventral extent of layer II. Layer II was initially divided into 10 subregions, which were recombined into three regions for this figure. A log scale has been used to emphasise differences in the proportion of high-intensity pixels.

3.4.3.1 *Laminar enrichment of genes can be automatically classified with high accuracy*

Classification using mean laminar expression I first looked to classify ISH images for each gene on the basis of mean expression across the three main layers. This required identifying thresholds for expression that would define enrichment. The correlation between mean expression across different layers is very high (Fig. 3.25 (a): II vs. III: $R^2 = 0.96, p < 10^{-5}$; IIvs.V : $R^2 = 0.88, p < 10^{-5}$; IIIvs.V : $R^2 = 0.90, p < 10^{-5}$). Although layer II shows slightly higher expression than layers III (Linear regression: slope = 0.88) and V/VI (slope = 0.98), the majority of

genes appear to be uniformly expressed. This suggests that variability in ISH labelling across different genes is higher than that across layers for a given gene.

This variability in ISH labelling intensity across genes precludes using absolute intensity level to identify the layers in which gene expression is particularly enriched. Instead, I normalised all mean pixel intensities to the sum across the layers, thereby providing a relative mean pixel intensity. Genes with high ISH labelling in a given layer are likely to be uniformly expressed across layers (relative mean between 0.3 and 0.4), whereas genes that are more weakly labelled in a given layer show an increased incidence of very differential expression (normalised mean values closer to 0 or 1) (Fig. 3.25(b)). This pattern is similar across layers, although very few genes show highest relative expression in layer III (Fig. 3.25(b)(ii)).

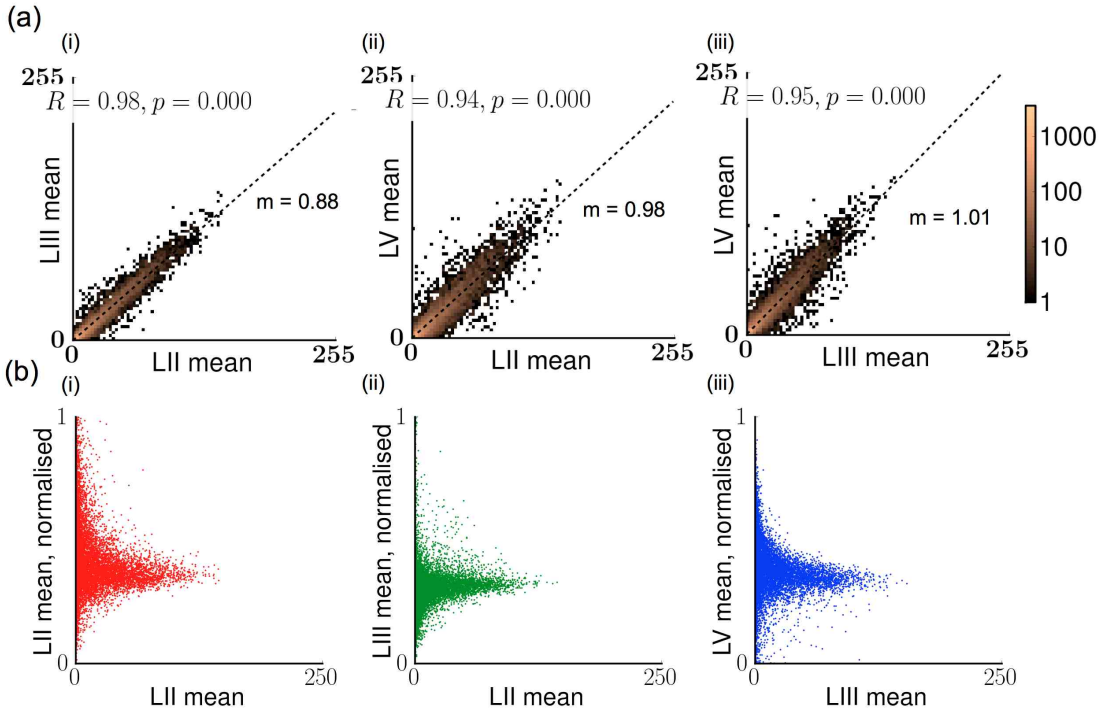


Figure 3.25: *The distribution of gene expression in MEC* (a) Two-dimensional histograms showing absolute mean pixel intensity plotted for LII region vs LIII region, LII vs LV/VI and LIII vs LV/VI. Individual layers means are highly predictive of other layer means (linear regression analysis, m is the gradient). Data shown using a log colour scale. (b) Scatter plots show the normalised means (relative to the other layers) plotted as a function of absolute mean for LII, LIII and LV/VI pixels. Data were normalised by dividing by the sum across the three layers. High intensity pixels are most likely to be observed when expression is uniform.

To set thresholds on relative mean pixel intensity that I could use to distinguish categories, I first examined the relative mean expression across layers of all the training data genes that had been manually validated. Relative mean expression clearly segregates genes specific to particular layers, and genes enriched in specific layers, but there is a high degree of overlap between enriched genes and uniform genes (Fig. 3.26(A)). I used the training data to approximate appropriate boundaries between classes (Fig. 3.26(B)) and optimised them by plotting a confusion matrix to show the proportion of training data genes in each group that were correctly automatically classified (Fig. 3.26(C)). A large proportion of manually determined

enriched genes are classified as uniformly expressed, whereas few uniformly expressed genes are incorrectly defined as being enriched. Thus, the classification is relatively specific, but not highly sensitive. Averaging accuracy across all the groups reveals that automatic classification using relative mean intensity has a sensitivity of 41.2% for detecting genes precisely, and of 53.1% for detecting genes as being enriched in the group they are actually enriched in or in a similar group. This means that 47% of enriched genes are automatically classified as uniform or as a different group. The specificity of this measure, measured simply as the proportion of uniform genes correctly detected as such, is 99.7%. This suggests that this classification scheme identifies only the most enriched genes, but misses many that appear visually enriched.

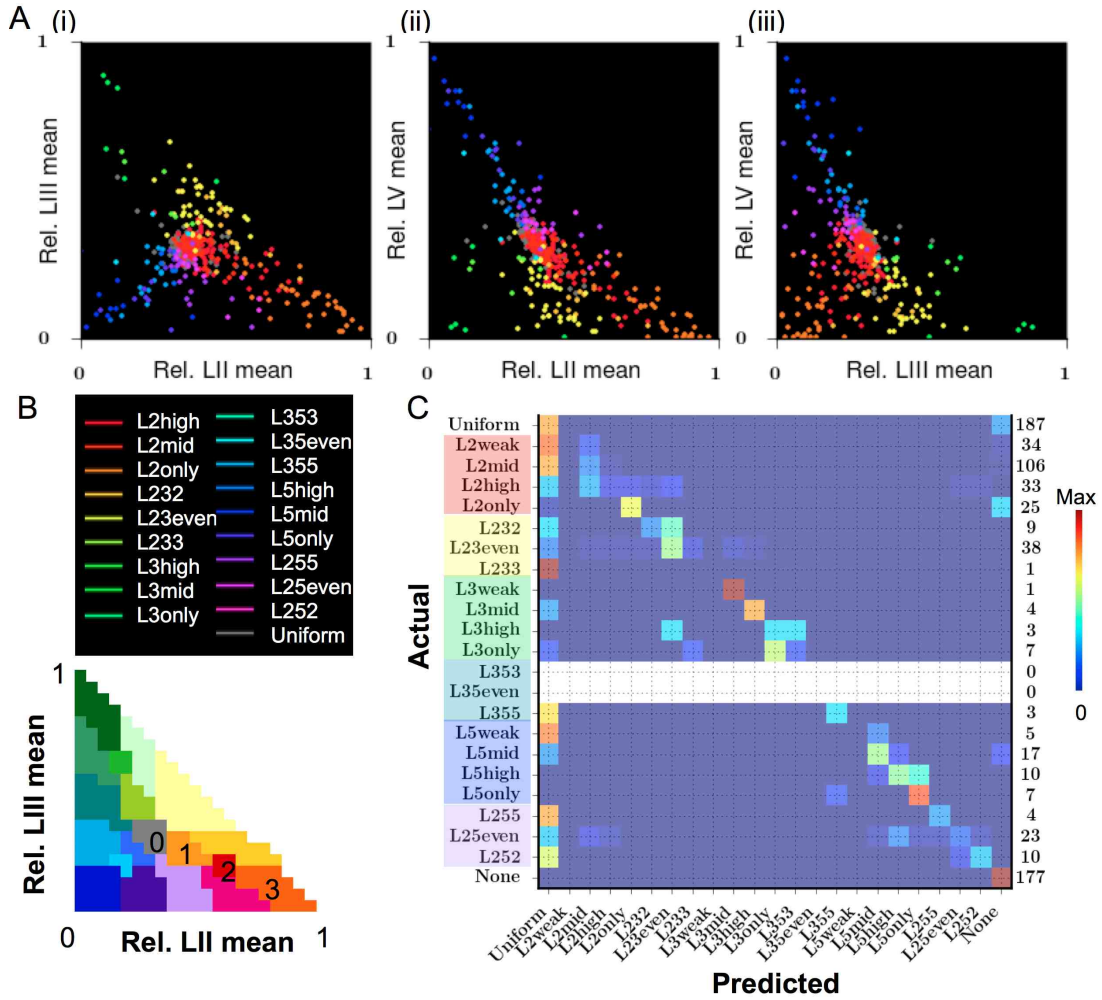


Figure 3.26: *Segregation of manually classified genes using relative laminar intensities.* (A) Scatter plots show the distribution of relative laminar mean intensities for training data genes (i) relative LIII intensity is plotted as a function of LII, (ii) LV vs. LII and (iii) LV vs. LIII. (B) Schematic showing how relative mean pixel intensity can be used to segregate groups to match the segregation that occurs in the training data (axes from 0-1). Numbers indicate different levels of specificity for genes enriched in layer II. (C) Confusion matrix showing the accuracy of classification for each group. Manual (actual) classifications are plotted against predicted classifications. The number of genes involved in indicated. Colours have been normalised such that each row sums to 1.

The disadvantages of using the mean are that it depends on the expression of all cells in a layer and does not take into account cell density. Visual assessment is potentially more sensitive to the proportion of cells in a region that show very high expression. Examples of images in which the mean is similar across the three layers, but the distribution differs are shown in Fig. 3.27 (A)(a-b). While Adcyapr and Eif4g2 appear to be enriched in layer II, Birc6 does not. The difference arises because the former genes show a high proportion of high-intensity pixels, particularly exceeding the mean (for the whole MEC region) added to 2 standard deviations, in layer II whereas the latter does not (Fig. 3.27 (A)(b)).

Classification using multiple features To address the question of which features to use for classification across all images, I examined the relationship between the relative mean expression, relative standard deviation in expression and the relative proportion of pixels in each layer exceeding specific thresholds, and assessed which provided the most information about laminar enrichment. Thresholds were calculated on an individual gene basis, using the pixel intensity distribution across the entire MEC region (see Fig. 3.25 (A)(b) for thresholds of example images). To make use of threshold data, I normalised it in the same way as the mean by calculating the relative proportion of pixels exceeding threshold in each region. A PCA analysis on the available features demonstrates that for layer II measures, the relative mean provides the most information about variability in the dataset, loading strongly onto the first component. The relative proportion of pixels exceeding the highest threshold loads strongly onto the second orthogonal component (Fig. 3.25 (B)(a)), which suggests that these features are important for explaining variability in the data. In addition, these principal components segregate genes with different levels of enrichment. The mid and high categories are not well segregated, but this may be because the difference was not well defined for the manual validation. For Layer V/VI all relative measures distinguish the enriched genes from uniform genes, which suggests that relative mean pixel intensity provides sufficient information (Fig. 3.25 (B)(b)). Absolute mean pixel intensity appears to further distinguish specific genes from enriched genes.

Plotting the relationship between the relative mean and relative proportion of high-intensity pixels across all genes reveals a highly positively correlated relationship that differs across the layers (Fig. 3.27 (C), grey). Many genes with moderate mean layer II pixel intensity relative to other layers show proportionally much higher high-intensity pixels in layer II than any other (Fig. 3.27 (A)(i)). This suggests that high-intensity pixels, and therefore high-expressing cells, are more likely to be found in layer II than other layers, which corresponds with results shown in Fig. 3.25 (A). However, high-intensity layer II pixels must often be accompanied by lower intensity pixels, thus reducing the overall mean. The use of these two measures enables good separation of the training data into classes (Fig. 3.27 (C), colours).

Of interest, the majority of genes manually classified as layer II (high or mid) actually show mean expression that is between 0.3 and 0.4 relative to the other two layers, suggesting relatively uniform expression, but they exhibit a much higher proportion of high-intensity pixels in layer II. There are two possible reasons for this: (1) layer II enriched genes are often enriched in just a subset of cells in layer II, with low expression in others, and (2) misalignment of layer II such that layer I pixels, which are low intensity so would reduce the

mean, are included as layer II. There are potentially mechanisms of differentiating between these two scenarios, but this analysis has not been performed.

I finalised list classifications by setting thresholds on the high-intensity pixels using similar methods to those described for the mean (Fig. 3.28(A)(a)) then combining the two methods so that only genes falling into select categories based on their relative mean expression are subsequently classified based on the distribution of their high-intensity pixels (Fig. 3.28(A)(b)). To increase confidence in this classification scheme I calculated a patterning score (PS) using the relative mean, proportion of high-intensity pixels and absolute difference in intensity between the highest expressing layer and others (Fig. 3.28(B)). This patterning score reflects how differentially genes are expressed, independent of the location of expression. It can be used to distinguish genes specific to one layer from genes that are mildly enriched, and to distinguish genes that are expressed at the same level across two layers compared with the third from those that are patterned to different degrees across all three layers (Fig. 3.28(C)). It can also be thresholded to dynamically adjust classification borders.

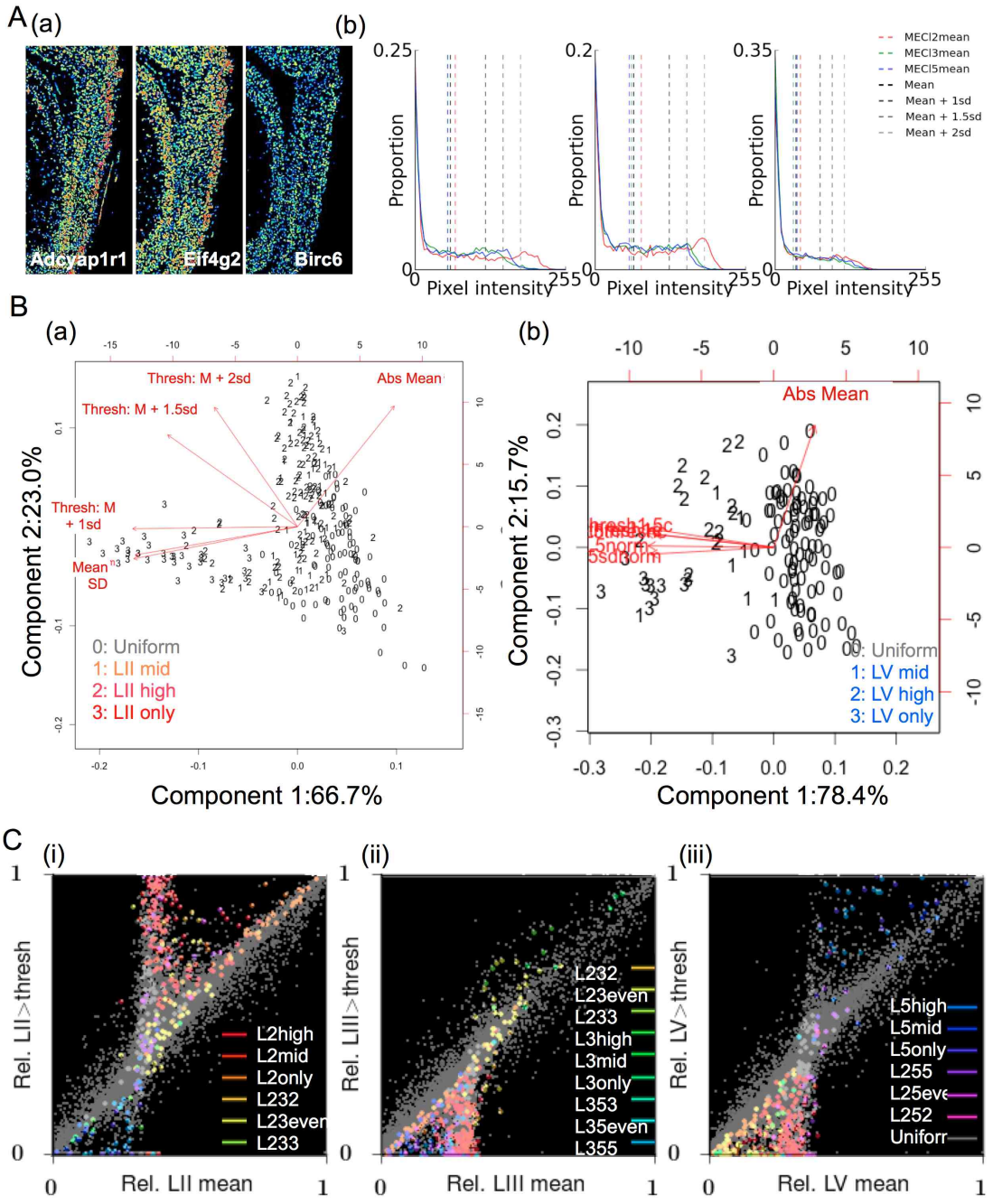


Figure 3.27: *Identifying features that distinguish enriched genes* (A) (a) Example images with a relative mean intensity in layer II of 0.3-0.35 but skewed distribution (b) Histograms showing the normalised distributions of pixel intensities in layers II (red), III (green) and V/VI (blue) for the images shown in (a). Vertical dotted lines mark different thresholds (see legend). Grey and black lines indicates thresholds set using data from the whole MEC region whereas coloured lines are layer-specific. (B) R biplots showing the distribution of individual data points as a function of the two principal components following a PCA analysis using data from the features shown. Numbers indicate the category that the data images were manually classified under. Red arrows indicate the loading of image features on the principal components (a) Layer II, (b) Layer V/VI. M = mean, sd = standard deviation. (C) The relative proportion of pixels exceeding threshold (mean +2sd) is plotted as a function of the normalised mean for pixels in (i) Layer II, (ii) III and (iii) V/VI. Data plotted in colours correspond to the data points within the manually validated training set whereas the grey region shows all data points from images in the ABA.

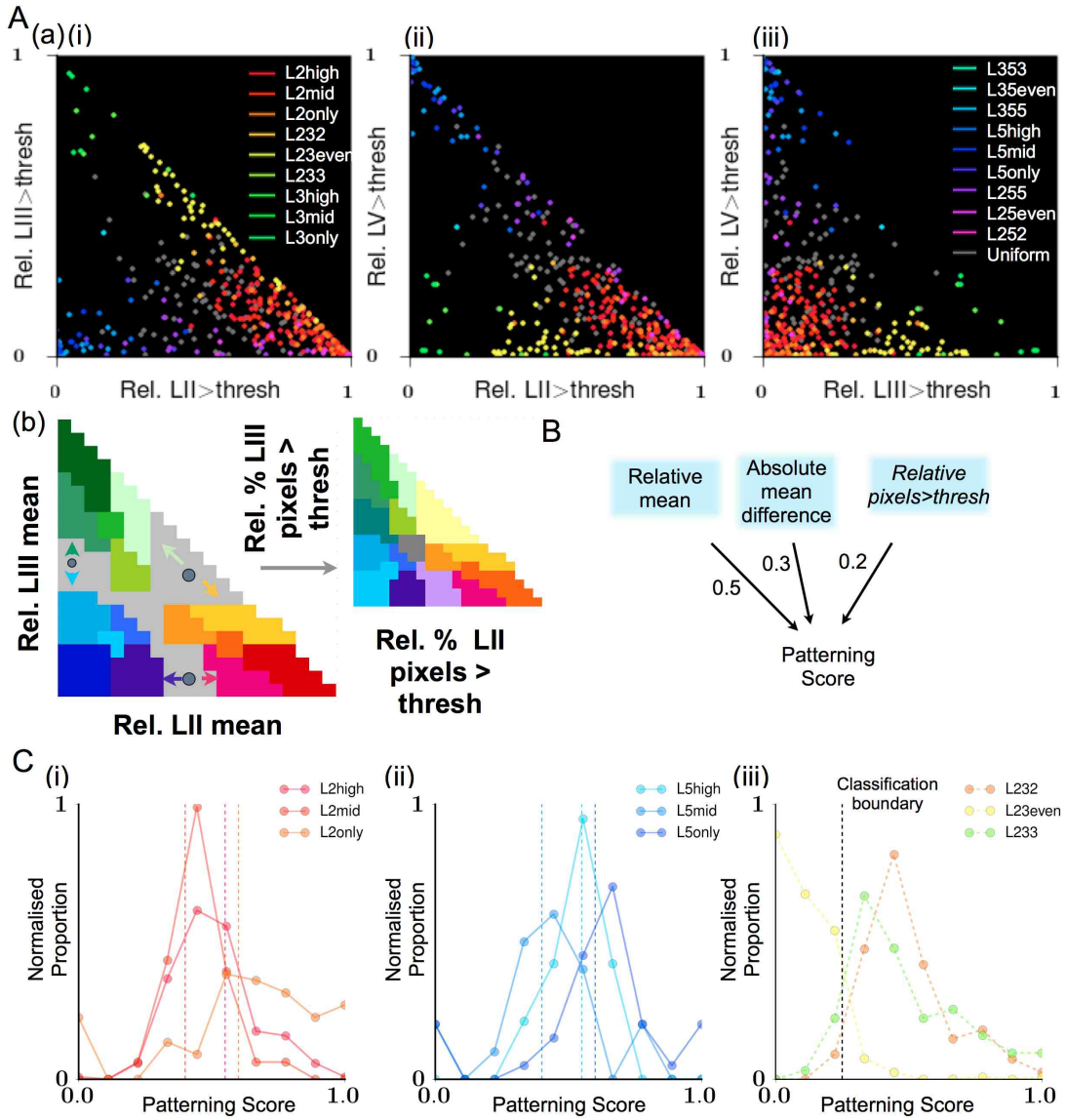


Figure 3.28: *Segregation of manually classified genes using optimised measures* (A) (a) Scatter plots show the relative proportion of pixels exceeding threshold in each layer as a function of the other layers for all images in each manually classified group. (b) Schematic showing how the combination of relative mean intensity (plotted from 0-1) and threshold measures (normalised across the layers from 0-1) was used to distinguish different layer classifications. Relative mean pixel values falling in coloured regions are directly categorised. Pixels falling in the 3 outer grey regions are re-classified into one of three surrounding regions depending on threshold scores. Pixels falling in the central light grey region are fully reclassified based on their threshold value. (B) Finalised procedure for calculating a patterning score (PS) for each image. The weighting of each feature is shown next to the arrow. (C) Demonstration of how effectively the PS distinguishes different levels of enrichment for (i-ii) enrichment of single layers, and (iii) differential enrichment in which two layers show enrichment of a gene compared with the third layer, but not necessarily to the same level. Vertical dotted lines indicate thresholds that act as class boundaries.

Genome-side semi automated classification reveals that layer II contains the highest proportion of enriched genes By combining a classification scheme that localises gene expression and a patterning score that determines the level of enrichment, I was able to classify genes into 23 categories (Fig. 3.29(A)) with 90% accuracy on average. Accuracy was assessed by visually

assessing genes in all groups except those labelled weak, uniform, or none. The sensitivity of this classification scheme is 90.2%, without loss of specificity at 99.2%. Thus, 90% of genes that are manually classified as enriched are also identified as enriched in the same category using automated classification. However, note for the calculation of specificity that only a subset of uniform genes have been manually validated.

This classification scheme is particularly accurate for layer II-enriched and layer V-enriched genes, whereas a number of genes manually classified as being enriched in layer III were predicted initially to be expressed in both layers II and III. This error resulted from very small misalignment of the sections. In total, over 2000 genes ($>$ 10%) are enriched to some degree in different layers. The advantage of having multiple groups for laminar enrichment (high, mid, weak) is that users can select which genes to include in the analysis depending on how sensitive or specific their requirements are. Sensitive analyses would include genes classified as 'mid' and 'weak', whereas specific analyses might only include the more enriched groups to ensure no inclusion of false positives.

The results of the NeuroBlast search suggested that layer II contains the highest proportion of differentially enriched genes (Fig. 3.23), but since the search was targeted towards enriched genes, there may have been a bias towards this layer. Analysis of the entire genome reveals that layer II does contain the highest proportion, at nearly 60% of all differentially expressed genes, if weakly enriched genes are included (Fig. 3.29(B). Only 3% of enriched genes are enriched in layer III over the other layers. Genes expressed in both layers III and V/VI, but not layer II, are very rare while genes expressed in layers II and III are very unlikely to show higher expression in layer III.

The distribution of relative mean pixel intensity and proportion of high-intensity pixels is shown for all manually validated (and uniform) classifications to show how effectively the groups can be segregated (Fig. 3.29(C)). Genes do not cluster into clearly distinct groups without overlap, but the measures do allow certain groups to be extracted. Plotting the average gene expression for each classification group indicates the differences between them (Fig. 3.30). Genes classified as being higher in layer II than other layers are both higher in layer II and show a predominantly dorso-ventral gradient, as do genes expressed uniquely in layer II. A similar pattern predominates for layer III genes. Layer V/VI genes, however, appear to be predominantly expressed in the more ventral half of MEC with a dorsal region that shows lower expression. There is some evidence that a layer Va can be identified by a more intense region near to the cell body-free layer IV. It follows that genes expressed in layers II and V appear to show a ventro-dorsal gradient in layer V/VI, combined with a dorso-ventral gradient in layer II.

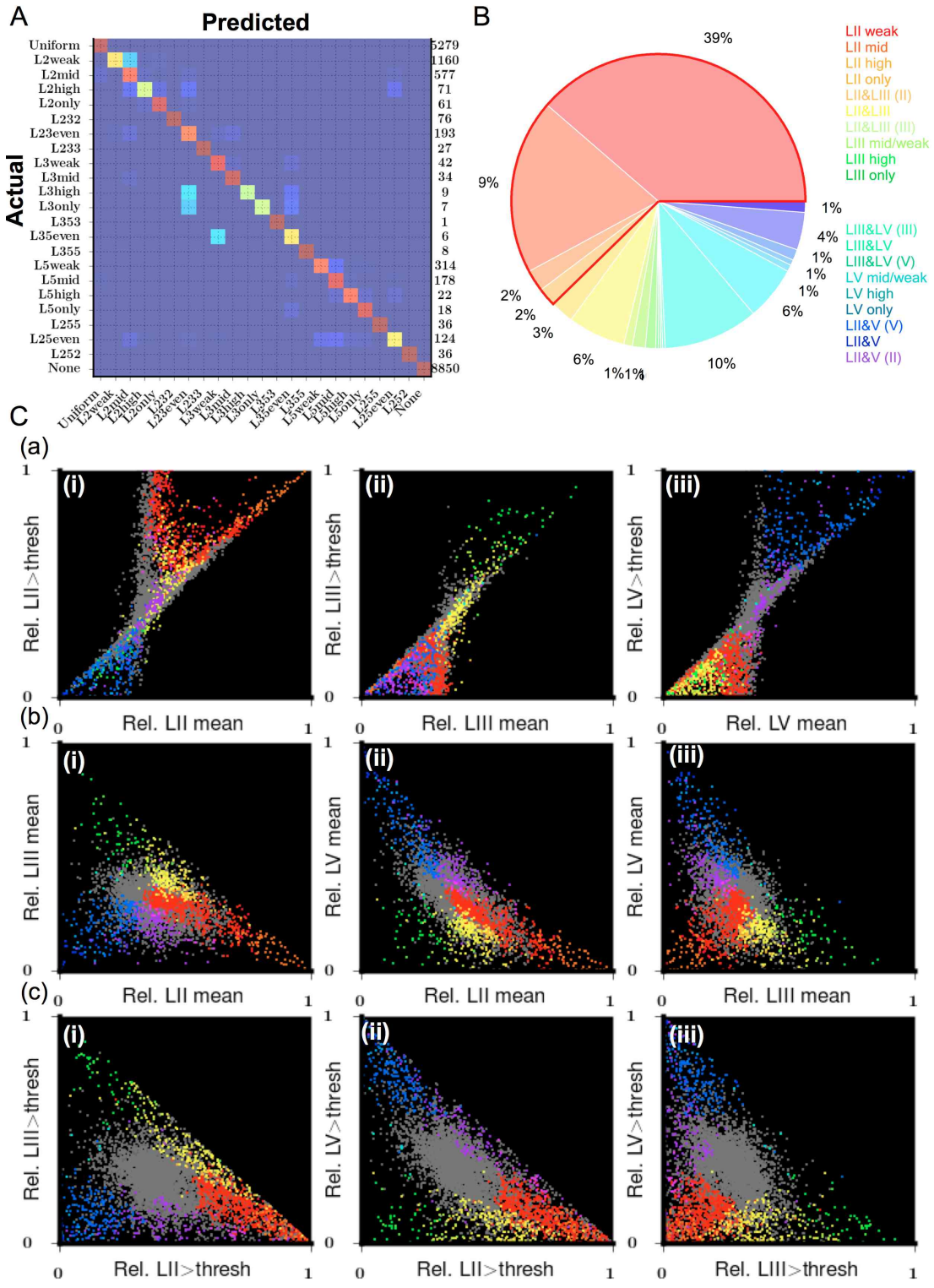


Figure 3.29: *The distribution of automatically classified layer-specific genes* (A) Confusion matrix showing the correspondence between the final gene classifications (after visual assessment) and automated categorisation. The number of genes in each category is shown on the right.Categories have been sorted so that neighbouring categories are more similar than distant categories (except LII high and LII & V/VI). Data have been normalised so that each row sums to 1. (B) Pie chart showing the distribution of enriched genes (excluding uniform and zero-expressing genes). Layer II expression is indicated by red, layer III by green and layer V by blue, with intermediate categories indicated by a colour mix. (C) Scatter plots show, for all final gene classifications: (a) the proportion of (i) LII, (ii) LIII and (iii) LV pixels exceeding threshold plotted as a function of the respective relative mean for each layer, (b) the segregation of the groups by relative mean in (i) layer II vs layer III, (ii) layer II vs layer V, and (iii) layer III vs layer V, and (c), as (b) for the proportion of pixels exceeding threshold.

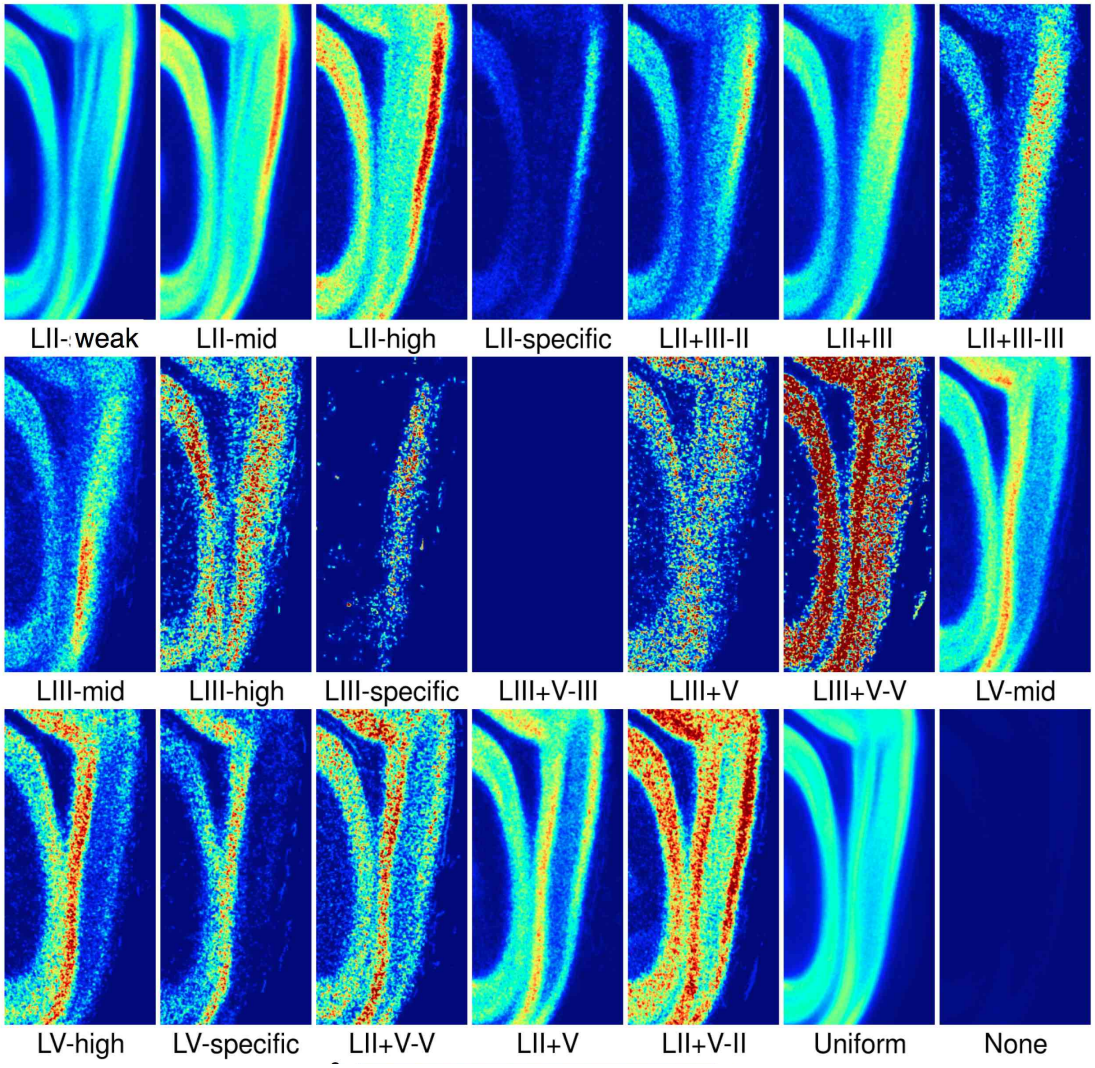


Figure 3.30: *Average gene expression patterns across layer-enriched genes* Images show average gene expression of genes in each laminar category in the 170x360 pixel region in which the MEC lies in the ML3 reference image. Every gene image that has been automatically classified in each category and manually validated (excl. LIIweak, uniform and none), and that satisfies quality control has been included. Pixel intensity is indicated by a colour scale from red (high) to blue (low). There were no suitable genes in the LIII&V (III) group.

3.4.3.2 *Classification of all ABA-databased genes across MEC confirms dominance of dorso-ventral gradients over ventro-dorsal gradients*

In order to classify genes based on the presence of a gradient, I segregated the MEC layers into 10 equidistant regions and calculated the linear regression slope relating mean or standard deviation in pixel intensity to distance from the dorsal border. I also calculated the relative proportion of high-intensity pixels in each region. For layer II and III pixels, the slopes across all the genes are predominantly zero or negative (dorso-ventral) (Fig. 3.31(A)(i)), whereas for layer V/VI pixels, a similar proportion of genes show no gradient, but a larger proportion show a positive (VD) gradient. This pattern holds whether the gradient in mean, standard deviation or proportion of high-intensity pixels is plotted (Fig. 3.31(A)(i-iii)).

How are manually validated MEC genes distributed across the dorso-ventral extent of MEC? To estimate the significance of a linear regression gradient to the classification of a gene as being dorso-ventrally organised, I plotted the gradients in different layers for the training data. Genes classified as showing gradients in layers II and III (the largest group), with a few exceptions, all exhibit slopes less than -2 in both layers, whereas genes that are classified as showing a gradient only in layer II tend to show gradients in layer III that are greater than -3, whether the standard deviation or mean pixel intensity is used for calculating a gradient (Fig. 3.31(B)). However, the boundaries cannot be as clearly defined as laminar classification boundaries and there are no obvious clusters. There is also a large degree of variation amongst genes showing gradients, so again I calculated a patterning score to reflect the degree of non-uniformity along the DV axis (Fig. 3.19(E)).

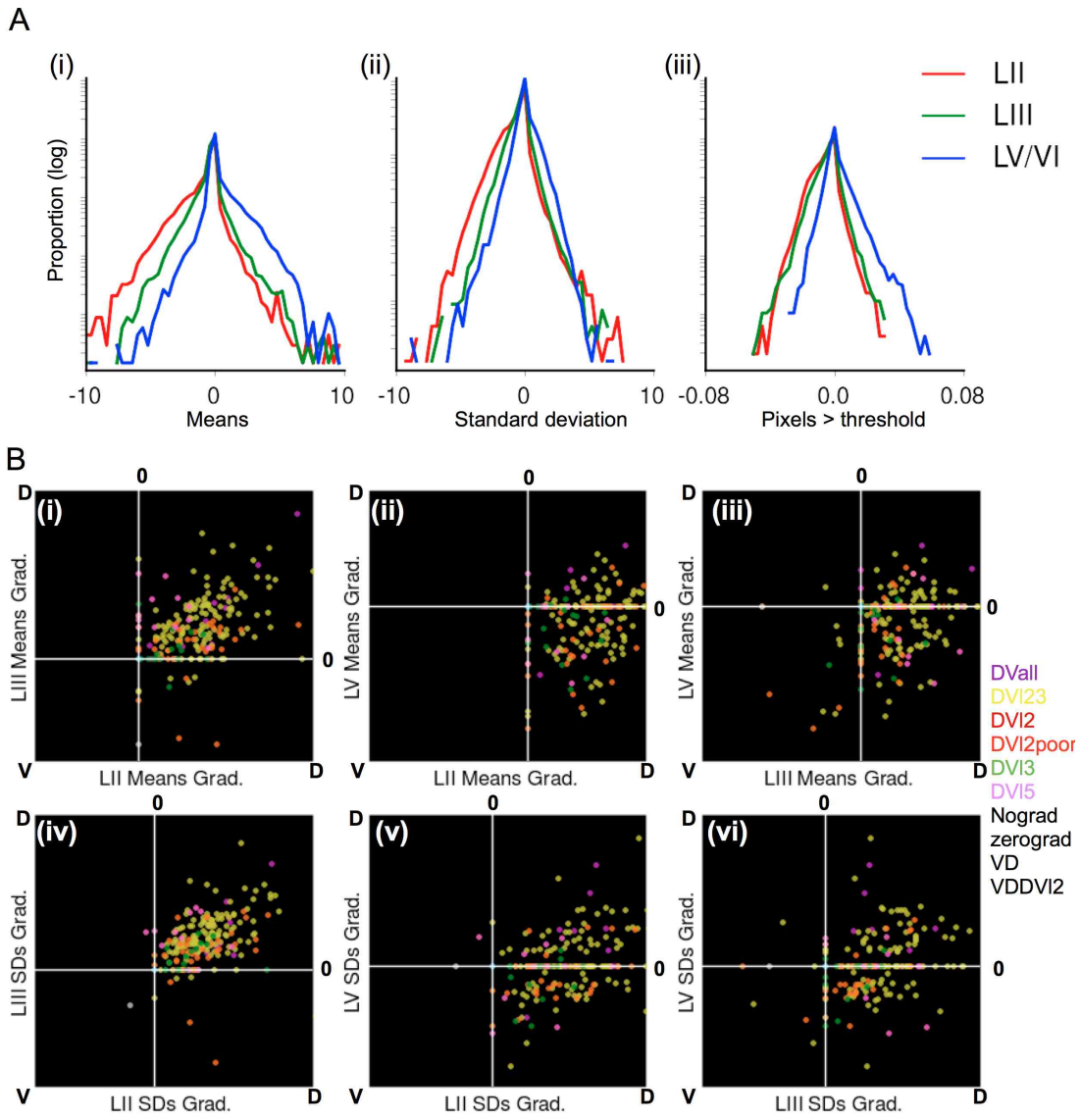


Figure 3.31: *Gradients in MEC* (A) Histograms show the distributions of gradients in layer II (red), layer III (green) and layer V/VI (blue) for differences in (i) mean pixel intensity, (ii) standard deviation, and (iii) threshold pixel intensity. Gradients are predominantly from dorsal to ventral (negative) in layer II, whereas they are positive in layer V/VI. The y axis has been plotted on a log scale. (B) Scatter plots show the gradients in (i-iii) mean pixel intensity and (iv-vi) standard deviation of manually classified genes in two different layers. Most genes show a DV gradient in (i,iv) layers II and III, whereas they tend to show similar and opposing gradients with similar frequency for (ii,v) layers II and V/VI, and (iii,vi) layers III and V/VI. D: Dorsal; V:Ventral; o: No gradient (at significance level of 0.3). Positive scores indicate DV gradients.

Final classifications: The difficulty in classifying genes based on gradients is reflected by the limited correspondence between some of the final manual classifications and the automated predictions (Fig. 3.32(A)). Nevertheless, some clear patterns emerge. The majority of genes show no gradient and these genes are rarely predicted to express a gradient, suggesting that the number of false positives (for the presence of a gradient) is low (Fig. 3.32(A-B)). A large proportion of these genes exhibit no gradient because they are not expressed in MEC

(labelled zerograd). The majority of other genes show a dorso-ventral gradient, with a very small number showing a ventro-dorsal gradient (Fig. 3.32(B)). The classification of genes with a DV gradient in layers II and/or III is relatively accurate (Fig. 3.32(A)), whereas genes showing a ventro-dorsal gradient are typically classified as showing it across all layers even when the gradient is more selective. This is likely due to poorer alignment at the ventral extent of MEC than the dorsal.

Most genes that are expressed with a gradient exhibit higher expression in dorsal layer II than ventral layer II (> 80%), of which a large proportion also exhibit a gradient in layer III (Fig. 3.32(B)). Genes expressed with the opposing gradient account for only 5% of genes with non-uniform DV expression. The differences between the groups is shown by the average images plotted in Fig. 3.33. Genes showing dorso-ventral gradients that include a DV gradient in layer II are highly expressed and show a gradient across all layers, particularly layers II, III and the inner region layer V, putative Va. Because the classification is not 100% sensitive, genes classified as dorso-ventral, but not in layer II (DV III and DV V), also appear to show a gradient in expression in layer II, but their overall expression is lower than the previous category. Genes showing no gradient in the dorso-ventral axis, on average, show relatively even expression across layers. The expression of genes showing a ventro-dorsal gradient across all layers appears to be high in ventro-medial MEC, a region that exchanges topographical connections with temporal hippocampus, and it gradually decreases more dorsally. Of interest, these genes show much lower expression, on average, in the visual cortex and subiculum, shown dorsal and anterior to MEC in these images, than dorso-ventral genes (Fig. 3.33).

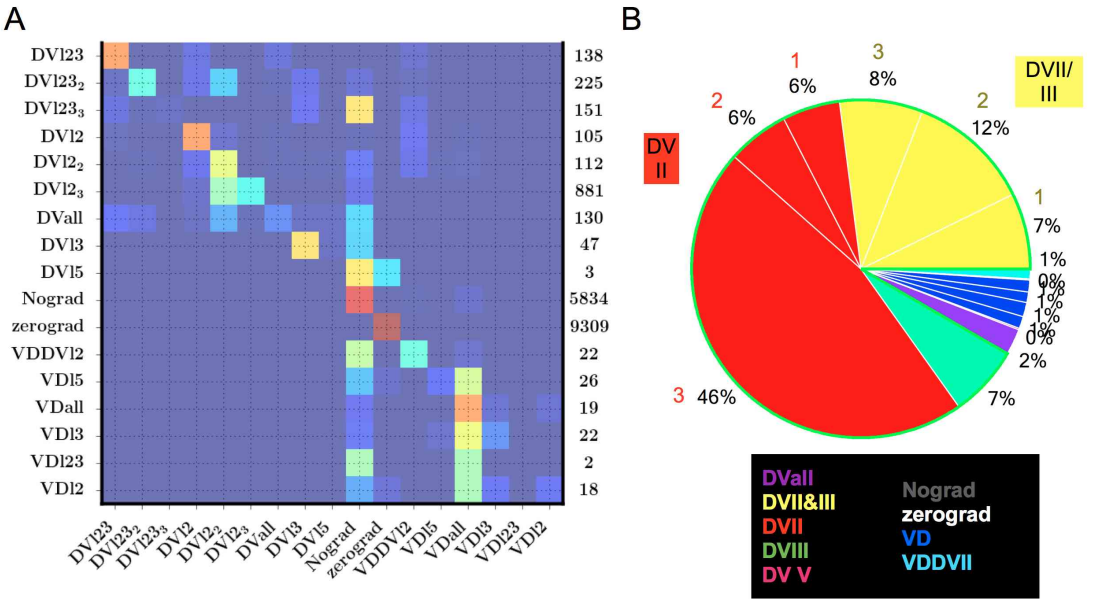


Figure 3.32: *The distribution of gradients in gene expression by MEC layer* (A) Confusion matrix showing the correspondence between the final gene classifications (after visual assessment) and automated categorisation. The number of genes in each category is shown on the right. Data have been normalised so that each row sums to 1. (B) Pie chart showing the relative numbers of genes with different types of gradient. Coloured numbers indicate different levels of specificity (see A).

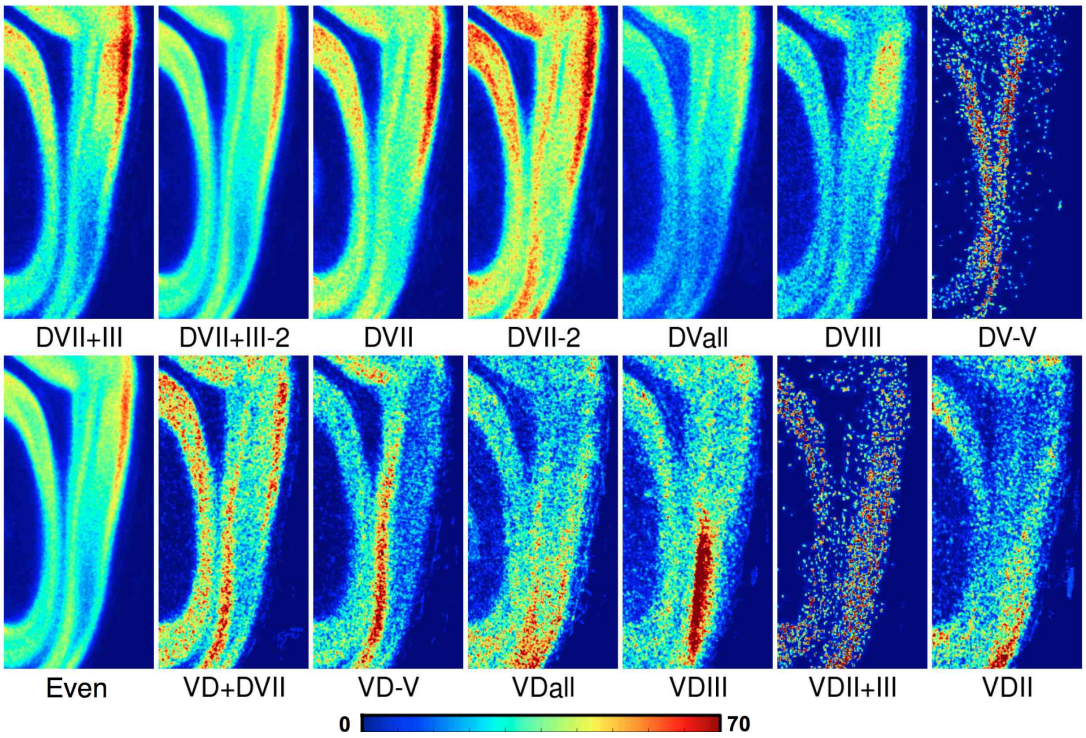


Figure 3.33: *Average expression patterns across different DV groups* Images show the mean gene expression for each category in the 170x360 pixel region in which the MEC lies in the ML3 reference image. Every gene image that has been automatically classified and manually verified in each DV category and that satisfies quality control has been included. Pixel intensity is indicated by a colour scale from red (pixel intensity /255) to blue (0).

3.4.3.3 *Gene expression patterns are consistent across the immediate medio-lateral extent*

All the classifications described have been defined based on expression in a single plane, the ML3 plane, or in the ML2 plane where an accurately registered image was not available for ML3. In order to investigate how consistent these classifications are across the medio-lateral extent of MEC, I plotted pixel intensity data for ML3 and compared it with ML2 (Fig. 3.34(A)(a)) and ML4(Fig. 3.34(A)(b)). Overall, relative mean pixel intensity appears to be consistent across the layers. The most considerable difference between ML regions is for layer V/VI: in ML4 pixel intensity appears to be lower for the majority of genes than in ML3, whereas ML3 and ML2 show a very similar pattern. This is consistent with the observation that many layer V/VI genes show lower expression in the most dorsal region, which is the only part of layer V/VI visible in ML4, than more ventrally.

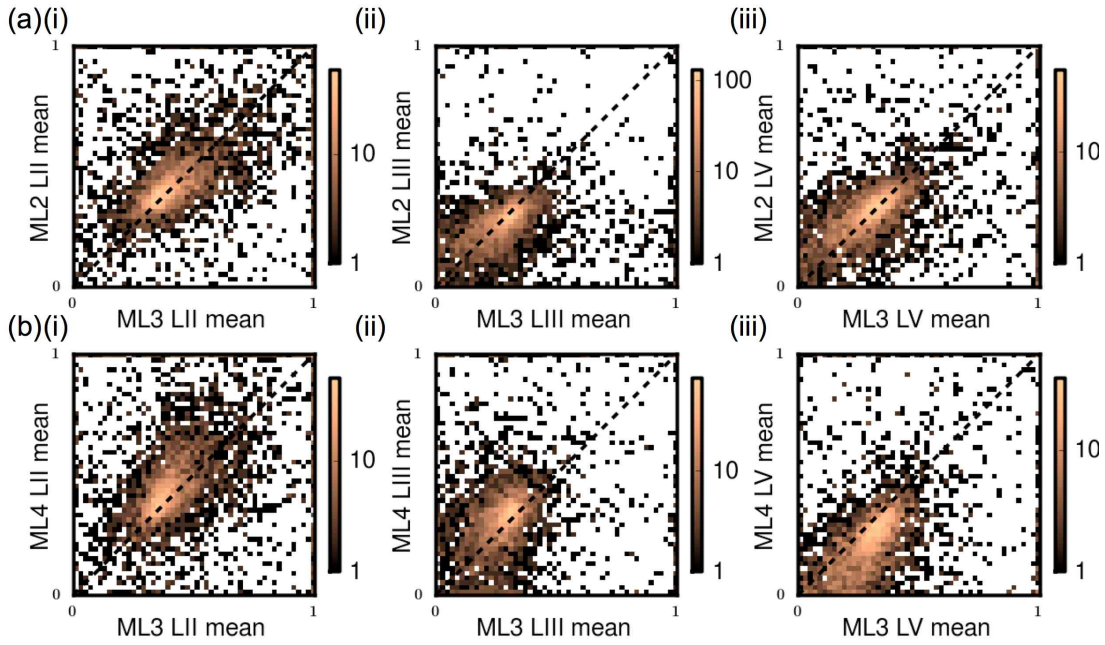


Figure 3.34: *Medio-lateral changes in gene expression* (A) Two-dimensional histograms showing relative mean pixel intensity in (i) layer II, (ii) layer III and (iii) layer V-VI of (a) ML2 images and (b) ML4 as a function of relative mean pixel intensity in the same layers for ML3 images. The colours are shown on a log scale to emphasise genes showing different normalised expression across the medio-lateral extent. The unity line is shown by the dotted black line.

3.5 DISCUSSION

In this chapter I investigated whether gene expression varies across MEC, and what types of patterns emerge when we examine differences in gene expression across the whole genome at the resolution of single cells. I show that gene expression is not uniform across MEC, but ranges from highly specific laminar expression to similar expression across layers. Moreover, a considerable proportion of genes are expressed with a dorso-ventral gradient, particularly in layer II, and commonly in layer III. Ventro-dorsal gradients that extend across MEC are far more rare, but layer V/VI expression often follows a ventro-dorsal gradient, possibly indicating that there are differences in expression between layer V and VI.

3.5.1 *Gene expression distribution and MEC anatomy*

Genes that show layer-specific enrichment are most likely to be expressed in layer II of MEC, a layer that send projections to the dentate gyrus of the hippocampus (van Groen et al., 2003), where axons may make contact and synapse onto newly-adult-born granule cells (Nakashiba et al., 2012). They show a prevailing gradient in expression from dorsal to ventral regions, which corresponds with physiological gradients (Hafting et al., 2005; Fyhn et al., 2008), and the morphological and synaptic integrative properties of individual cells (Garden et al., 2008). Regions of MEC project with a topographic organisation to the dentate gyrus and CA3, such that dorsal regions target septal areas, and ventral regions target temporal areas (van Groen et al., 2003; Dolorfo and Amaral, 1998b). Since CA3 has been shown to exhibit gene expression gradients along this axis (Thompson et al., 2008), the dorso- ventral gradients in MEC may map onto these.

While gradients in gene expression in the superficial layers of MEC are predominantly dorso-ventral, gene expression does not show this pattern to the same extent in the deep layers. MEC layer V and VI receives the major projection from the CA1 region of the hippocampus (Swanson and Cowan, 1977), which is also topographic (Tamamaki and Nojyo, 1995). CA1 also shows a septo-temporal organisation in gene expression (Dong et al., 2009), so it is not clear how this relates to gene expression patterns in layer V and VI of MEC.

Genes are commonly expressed in layers II and V and in layers II and III, but it is very rare to find genes expressed at higher levels in layers III and V than layer II. Does this imply that there are few molecule-specific mechanisms in layers III and V that do not also occur in layer II cells? All layers of MEC contain pyramidal cells, but many other cell types co-exist, and have not been well characterised. The predominant cell type in layer II is the stellate cell, whereas layer III contains fewer (Gatome et al., 2010; Canto et al., 2008). Layer V also contains horizontal and polymorphic neurons and all the layers contain a number of multipolar projection neurons and interneurons (summarised in Canto et al. (2008)). It is possible that differences in cell type and in the types of inputs the cells receive explains this asymmetry.

I observed in layer II that a considerable proportion of genes appear to be expressed at high levels, but only in a subset of cells. This pattern was observed for a small proportion of genes in layer V/VI, but not to the same extent. One possible reason for this could be simply

that the alignment of layer II is incorrect such that low intensity layer I pixels are included in the layer II analysis. An alternative possibility is that genes are more likely to be enriched in a subset of cells in layer II than in the other layers. It will be important to test this more thoroughly, as cell-type specific indicators can be very useful for elucidating specific neuronal functions.

3.5.2 *Using ISH data to extract gene expression patterns*

I have extracted gene expression data using purely *in situ hybridisation* data collected for the Allen Brain Atlas and quantified it based on relative staining patterns across regions. This data is very informative, not only because of its spatial resolution, but the fact that regions over a very large area can be easily directly compared. Because of its spatial resolution, ISH can provide information on laminar expression, which is of particular interest here, as well as cell-type specific expression patterns if the distribution of cell types is reasonably stereotyped. In both cerebellum (Kirsch et al., 2012) and CA1 (Dong et al., 2009), cell-type specific expression information has been extracted from ISH data.

Of course, there are several caveats to the use of ISH. The principal issue perhaps is that the expression of most genes is only measured in one mouse and it is not measured in the same experiment as the expression of most other genes. This means that ISH images like those in the ABA provide no cell type information unless the specific organisation of a cell type is already known. This is an important problem, and certainly limits the conclusions that can be drawn, but the fact that some genes show laminar expression while others don't remains interesting, particularly since some patterns are highly consistent across many genes. A second problem with the nature of this analysis is that ISH is not directly quantitative and it is very difficult to compare level of gene expression when probe efficiency plays an important role in determining level of staining. I have made the assumption in this work that probe binding efficacy is relatively equivalent across different brain regions, but this is not necessarily the case. There is no way out this, but the efforts of the Allen Institute to make the experiments as similar as possible across all genes, and the fact that comparisons for individual genes are being made based on relative rather than absolute expression means that this ISH information can be used in a semi-quantitative manner. Furthermore, the ISH experiments in the ABA have been validated for the top 1000 queried genes, suggesting that their staining patterns are typically representative.

It is also important to note that ISH provides a level of detail at a completely exploratory level that is unavailable to other tools. It reveals the sparsity or density of expression within a region of interest, and it can reveal previously unknown gradients in expression along any direction of axis. It also provides a means of analysing the relative expression of a gene across very large regions, at the resolution of single cells. Analysis by microarrays and RNA-seq can also have very high spatial precision if RNA data is collected by specific cell types, but this requires more prior knowledge. The major disadvantage of the use of ISH is that the results are highly dependent on the probes used, a limitation overcome by the use of sequencing.

3.5.3 *Using the Allen Brain Atlas tools to extract gene expression patterns*

I explored using tools from the Allen Brain Atlas to identify enriched genes, particularly when this project was in its infancy. They are demonstrably powerful for identifying examples of genes with strong enrichment patterns (Fig. 3.27(C)). However, it is very difficult to quantify the likelihood that all patterns of interest have been identified. The correlation score, which is the only measure of similarity provided for the NeuroBlast, gives limited indication of the likelihood that a gene will be correctly classified. However, its accuracy is limited and the number of genes with very high correlations is low, so genes with lower correlations must be included in the search. There is therefore no measure that accurately predicts how reliable NeuroBlast will be at predicting similarly expressed genes, which limits its use as a tool for identifying enriched genes.

3.5.4 *A resource that can be used to probe gene expression patterns throughout the brain*

I have developed a resource that provides a means of querying gene expression at pixel resolution in any region of interest in ISH images from the ABA. I have described the use of this tool to generate lists of enriched genes, in this case within MEC, that can be assessed for accuracy and flexibly adjusted depending on the level of sensitivity and specificity required. I could generate such a list for any other brain region or combination of brain regions. This makes the resource highly flexible and useful for a number of different types of question. For example, gene expression may be informative about functional similarities and differences between different brain regions. Such an analysis could also establish genes with high regional specificity that are useful as systems neuroscience tools for isolating control of specific cells. The benefits of this particular style of analysis are as follows:

- *Easy to expand:* The procedure I have used for processing and extracting data from the images can be used for any sagittal or coronal section in the brain, although registration and segmentation techniques would need to be individually optimised. It can not only be used to identify enriched genes but to compare different brain regions by the genes they express.
- *High resolution:* The 2D images from the ABA have $\sim 10\mu m$ resolution in the sagittal plane, so it is possible to investigate very local patterns of expression. This resolution could even be increased given sufficient computing resources. I have shown examples of highly specific expression patterns within MEC, but such expression may also be identifiable in regions of neocortex, throughout the neocortex, or in non-cortical regions. This resolution is considerably higher than that used in the 3D ABA analyses.
- *Relatively easy to estimate the accuracy of the method:* Because it is straightforward to directly relate the data extracted to its original image, there are many possible methods for identifying errors. No method is perfect, but it is likely that several used together can detect the majority of images with errors. With the ABA volumetric analysis it is much more difficult to quantify the degree of error for each gene.

3.5.4.1 *Limitations of the resource*

The most significant limitation of this resource is that it extracts patterns from registered images that have not been 100% verified as being of high quality. However, I developed the resource such that a number of different types of error could be detected, and I set up the classification procedure so that classifications could be made on the basis of higher or lower sensitivity to errors, thereby increasing its robustness. At one stage throughout the classification process, I investigated whether lack of correspondence between visual assessment and automated classification may be due to poor visual assessment. For example, it is possible that the eye is not highly sensitive to detecting certain types of gradient. This would make the validity of optimising classification to match the manual validation of the training data uncertain. In theory, if images have been accurately registered, then the automated classification could be considered a 'ground truth' for the presence of a gradient since it directly reads pixel data from the images. However, some imperfections in the expression images, such as the pial surface being detected as a cell, as well as slight misalignment in registration, could not easily be automatically detected, even after error detection steps. As such, comparing automated classifications to manual classifications provided an estimate about the level of error that was obtained through automated classification.

The resource can be used to identify genes that are expressed with a gradient, in this case the dorso-ventral axis. One potential caveat of this is that changes in cell density could either cause differences in gene expression, for example a dorso-ventral gradient, or mask them, in the case of a ventro-dorsal gradient. Research in which cell densities have been calculated in MEC using Nissl staining suggests that there is no difference in the number of dorsal and ventral stellate or pyramidal cells in mouse MEC (Gatome et al., 2010), although a gradient in cell density has not been explicitly tested. However, there is evidence that somatic size is larger in dorsal regions (Gatome et al., 2010; Garden et al., 2008), which could impact on the measure used. The most effective method to ensure that a dorso-ventral gradient exists would be to ensure that the proportion of high-intensity pixels changes more from dorsal to ventral than would be expected from the change in somatic area of the cells. It might also be possible to normalise gene expression based on the average expression of all genes or using a marker such as Nissl to quantify the area occupied by cells across different regions. In this work I have only classified genes as dorso-ventral if they have a strong gradient, a high patterning score, which takes into account the proportion of high intensity pixels, and they have been manually checked. As such, it is unlikely that genes with expression categorised as following a gradient result simply from structural differences. Furthermore, the analysis used here can detect many genes that show no difference in expression level along the gradient, as well as genes expressed with a ventro-dorsal gradient.

3.5.5 *Conclusions*

Despite limitations, the ISH data and 2D image registration resource have enabled me to provide a molecular profile of the MEC that reveals that a small but significant proportion of genes are enriched in particular layers of MEC and they are often dorso-ventrally organised.

This approach could be directly applied to any other region of the brain in the sagittal sections registered so far, and could also be extended to higher resolution, different sections and for different questions.

3.5.6 *Future work*

In this chapter, I have examined MEC gene expression at a global, data-driven level. I have not identified the genes that show particular patterns of expression or determined how their functions might relate to known MEC physiology. I also have not considered how MEC expression relates to other brain regions. In the next chapter I will address these topics, as well as looking more closely at the anatomical patterns found in gene expression here to determine whether the MEC has a genomic anatomy beyond our current understanding.

4

GENE EXPRESSION AND FUNCTION IN MEDIAL ENTORHINAL CORTEX

4.1 SUMMARY

In this chapter I use the resource outlined in the previous chapter to answer several questions about the genetic profile of MEC: (1) Is MEC genetically distinct from other regions of neocortex with different functions? (2) What types of genes show enriched expression in MEC? (3) Are genes enriched in different MEC layers functionally distinct? (4) Are MEC layers differentially associated with pathology, (5) Are genes with dorso-ventral gradients functionally distinct? (6) Are dorso-ventral genes associated with pathology? (7) What other patterns of expression can be identified in *in situ hybridisation* (ISH) data and can this inform us about the anatomy and physiology of MEC? These questions rely on combined use of the Allen Brain Atlas dataset and tools developed by the ABA and myself that provide gene expression information for the entire MEC at the resolution of approximately 10µm.

1. *Is MEC genetically distinct from other regions of neocortex with different functions?* The relationship between gene expression patterns in the MEC and other neocortical areas processing visual or somatosensory (SS) information is weaker than that between visual and somatosensory areas. A small number of genes (198) show more than 60% of their summed expression in the three regions in MEC rather than in visual or SS cortices.

2. *What types of genes show enriched expression in MEC?* There is little evidence that any functional component is enriched amongst MEC-enriched genes. There is evidence that genes involved in protein localisation and certain signalling pathways are overrepresented, but this is not robust. MEC-enriched genes show typical evolutionary conservation of genes included in the Allen Brain Atlas and show no increased likelihood of being involved in disease pathways.

3. *Are functionally relevant genes differentially distributed?* Genes with certain functional annotations are differentially overrepresented amongst layer-enriched genes. While plasticity-related genes, axon guidance-related genes, mitochondrial genes and ligand-gated receptors are most likely to be enriched in layer II, transcription factor activity, glutamate receptors and MAPK-signalling related genes are most likely enriched in layer V/VI. Few genes are specific to layer III but all layer-specific genes are enriched for calcium ion binding-related genes and dendritic genes, which suggests that genes within these families, which likely play an important role in controlling input-output relations in the cell, could contribute to differences in layer functions.

4. *Are MEC layers differentially associated with pathology-related genes ?* Layer II genes are very highly likely to be involved in molecular pathways that are disrupted or targeted in Alzheimer's, Parkinson's and Huntington's disease, accounting for almost one quarter of the genes involved. This suggests that layer II may be the most vulnerable to neurodegeneration, which is consistent with the extensive pathology found there (Braak and Braak, 1991). Layer II genes also show a high proportion of epilepsy and autism-related genes, whereas schizophrenia-related genes are more overrepresented amongst layer V-specific genes.

5. *Are genes with dorso-ventral gradients functionally distinct?* Mitochondrial genes and cytoskeletal genes all distinguish dorso-ventral from ventro-dorsal genes, whereas synaptic genes are equally enriched across genes with different types of gradient. Dorso-ventral genes can be distinguished from genes with no gradient and ventro-dorsal genes by the enrichment of intracellular signalling cascade genes, ion channel activity, axon guidance genes and MAPK signalling-related genes. Of particular relevance to the function of MEC in spatial memory, plasticity-related genes are overrepresented both in dorso-ventral and genes without a gradient in expression, but are underrepresented amongst ventro-dorsal genes. Immediate early genes, including the plasticity-related gene *Arc*, are also predominantly expressed with a dorso-ventral pattern, which may provide insight into differences in activity levels and types of plasticity throughout the extent of MEC. Ventro-dorsal genes appear to show a particularly increased likelihood of being involved in synaptic vesicle processes.

6. *Are dorso-ventral genes associated with pathology?* Genes expressed with a superficial layer dorso-ventral gradient are extremely enriched in lists of genes associated with neurodegenerative disease pathways, including Alzheimer's disease, Huntington's disease and Parkinson's disease. Enrichment is strongest for genes exhibiting enriched expression in layer II. Ventro-dorsal genes are much more likely than dorso-ventral genes to be involved in autism, and show similar enrichment for epilepsy and schizophrenia.

7. *What other patterns of expression can be identified in in situ hybridisation (ISH) data and can this inform us about the anatomy and physiology of MEC layers?* ISH images reveal genes that characterise very narrow layers of MEC, including putative layer Va and layer VI(b). Many of these genes are expressed in both these layers but not dorsal layer V/VI proper, and they are weakly overrepresented for post-synaptic density genes. Gene expression also distinguishes the dorsal border of MEC from an ambiguous region between MEC and the postrhinal cortex, which is putatively parasubiculum. In layer II there are genes that selectively mark patches of cells that invade the layer I region and appear to be clustered together. The calcium-binding protein calbindin is one such gene, and there is evidence that it is expressed in cells that project outside the hippocampus (Varga et al., 2010). 'Patchy' genes show moderately increased likelihood of being involved in synaptic transmission. The expression of different post-synaptic density genes also appears to distinguish dorso-ventral layer II genes from genes that are absent in the very most dorsal region, but show expression in intermediate regions, and genes that show

ventro-dorsal layer II expression, suggesting that the PSD variability is an important determinant of differences along the layer II gradient.

4.2 INTRODUCTION

The medial entorhinal cortex (MEC) provides an intriguing model system for studying the relationship between gene expression and cell identity and function. In the previous chapter, I outlined a resource that I have developed that, through analysis of high-throughput *in situ hybridisation* data in the Allen Brain Atlas, enables detailed, high-resolution analysis of gene expression patterns in MEC. Here I will present reasons for hypothesising that the laminar and dorso-ventral patterns of gene expression will be functionally organised and show how consistent these predictions are with ISH data from a large proportion of the mouse genome.

The MEC is a cortical region that primarily receives inputs from other cortical regions, the thalamus and limbic structures such as the hippocampus and amygdala. The projections of its cells terminate in similar regions (Insausti et al., 1997). Although it shares the laminar organisation of the neocortex, the organisation differs: the superficial layers primarily send projections to the hippocampus (Steward and Scoville, 1976; van Groen et al., 2003), as well as other regions, while the deep layers receive the major excitatory input from the hippocampal formation (Swanson and Cowan, 1977; Tamamaki and Nojyo, 1995). Many cells in MEC receive inputs spanning across multiple layers (Canto et al., 2012) and the superficial excitatory cells do not appear to share the intraconnectivity of the neocortex (Pastoll et al., 2013; Couey et al., 2013). These observations might suggest that the laminar organisation of gene expression differs between MEC and neocortex.

Another feature of MEC that leads to hypotheses concerning gene expression enrichment is that it is a component of the spatial navigation and learning system (see Moser et al., 2008, for a review). Lesions to MEC disrupt place field stability in CA1 (Brun et al., 2008a) and stimulation of MEC in humans may enhance learning of spatial location (Suthana et al., 2012). Moreover, grid fields in MEC rescale with changes to the environment (Barry et al., 2007). These results suggest that learning is an important feature of MEC networks. Further to this, the MEC appears to be particularly vulnerable to pathology in Alzheimer's disease (Braak and Braak, 1991; Devanand et al., 2007), which suggests that its cells are likely to express genes involved in the dysregulated pathways associated with the disease.

It is also possible to predict that different layers and regions within MEC may express different types of gene. In layer II there are two major types of projection neuron, stellate and pyramidal cells(Steward and Scoville, 1976), that have both been shown to function as grid cells (Domnisoru et al., 2013). A subset of layer II projection neurons projects to the dentate gyrus (Varga et al., 2010), one of the only sites in the brain where functional neurogenesis occurs (van Praag et al., 2002), so their genetic profile may reflect their requirement for the formation and maintenance of new synapses. Although it has been shown that this population express the gene *Reln* (reelin) (Varga et al., 2010), they have not been convincingly identified as being of a particular cell type. A second subgroup of layer II cells express that calcium-binding protein calbindin (Varga et al., 2010), but their function is not clear.

Layer III projects directly to the CA1 pyramidal cells and it is these inputs that when disrupted, affect place cell stability in CA1 but not in CA3 (Brun et al., 2008a). Layer III neurons also appear to be particularly sensitive to degeneration in human epilepsy (Du et al., 1993). Both the superficial layers contain pyramidal and stellate-type excitatory cells and

multiple different types of interneuron (Canto et al., 2008). Dendrites from both branch in layer I (summarised in Canto et al., 2008), suggesting that they can potentially receive many similar inputs, but may perform different computations, thereby providing different inputs to the hippocampus. While many deep layer neuron dendrites also extend into the superficial layers (Hamam et al., 2000), the projection targets of layer V/VI neurons are predominantly to the neocortex, rather than into the hippocampal formation (Insausti et al., 1997). Layer V cells show a unique electrophysiological profile that has been proposed as a neural basis for working memory (WM), particularly as it is robust to distractors (Egorov et al., 2002; Fransén et al., 2004), and it can be maintained *in vitro* in the absence of synaptic transmission, which suggests that it is an intrinsic property. These properties have been linked to calcium signalling (Zhang et al., 2011).

Although layers II, III and V are the major most well-studied cell body layers, cell types with slightly different properties exist in layer I and on the superficial border of layer V with the laminar dissecans, so-called layer Va. It is also possible to distinguish a layer VI between layer V and the white matter (summarised in Canto et al., 2008). Layer VI neurons also receive hippocampal and subicular inputs but unlike layer V neurons, only a subset of pyramidal neuron dendrites extend into superficial layers (Canto et al., 2008). They are often intermingled with layer V neurons and the white matter, so can be difficult to identify. Layer VI contains a population of neurons that are particularly sensitive to nicotine (Tu et al., 2009). Dorsal to layers II and III there also exists a cell population that is sometimes considered an extension of MEC, and have also been labelled as putative parasubiculum (Boccara et al., 2010; Burgalossi et al., 2011), but they have not yet been convincingly distinguished. These differences across layers and regions show there is clearly potential for identifying specific patterns of genes that characterise individual layers.

Aside from laminar differences, the MEC also exhibits variability in the spatial resolution of cellular physiological properties along the dorso-ventral axis of MEC (Hafting et al., 2008; Stensola et al., 2012). This feature could depend on the differential expression of certain excitability- and transmission-related genes and may interact with other topographically organised features such as the connectivity of cells (Dolorfo and Amaral, 1998b). Layer II stellate cells are known to exhibit a gradient in their synaptic integrative properties, and in the distribution of certain ion channels (Garden et al., 2008), which may be maintained by genetic gradients.

Together, these observations provide us with both specific predictions and a number of open questions that an understanding of patterns in gene expression, analysed using high-throughput ISH data, may aid in answering. In this chapter the goal is to understand which genes and patterns of expression distinguish the MEC from other regions of cortex, how MEC layers differ from one another and what other patterns of anatomical and functional interest gene expression highlights. The aims fall into three categories: (1) consideration of the attributes of proteins encoded by patterned genes, (2) establishing whether particular regions of MEC are differentially enriched for brain disease-related genes, and (3) using patterns in gene expression to further distinguish anatomical borders and cell types.

4.2.1 *Aims*

1. Identify differences in gene expression between MEC and neocortex.
    - to determine whether gene expression is more similar between neocortical regions than between MEC and neocortex.
    - to identify genes that are enriched in MEC relative to neocortex and determine whether their gene products are enriched for particular functional categories.
    - to establish whether MEC-enriched genes show an increased likelihood of being recently evolved.
    - to establish whether genes that are enriched in MEC have disease associations, and whether these map on to pathology in MEC.
2. Analysis of functionally-relevant attributes of genes that distinguish the major layers of MEC.
    - To establish whether particular gene attributes (cellular components, molecular functions, pathway involvement) are enriched amongst genes expressed in different layers of MEC.
    - to establish whether different layers in MEC show different probabilities of expressing genes that are more recently evolved, and how this relates to the functional properties of the layer.
    - to establish layer-specific enrichment of disease annotations, and map these on to pathology and layer function.
3. Characterisation of the dorso-ventral genetic profile of MEC
    - to explore whether genes expressed with a dorso-ventral gradient are associated with particular functionally-relevant attributes and whether disease pathway-related genes are more associated with some patterns of expression than others.
    - to investigate whether genes exist that show gradients that correspond to the modular gradients in grid spacing properties, and whether these genes fall into particular functionally-relevant groups.
    - to find out whether genes expressed with septo-temporal gradients in hippocampus exhibit dorso-ventral patterns in MEC.
4. Investigation of highly specific patterns of gene expression within MEC
    - to determine if ISH images can be used to define regions that are cytoarchitecturally difficult to identify in sagittal sections, such as layer VI, layer Va and Vb, the large patches dorsal to layer II, and whether certain cell types can be isolated.
    - to establish whether the genes that are uniquely expressed in particular cell types and regions within MEC have functionally relevant attributes that relate to physiological differences.

4.3 METHODS

4.3.1 *Gene expression patterns beyond MEC laminar enrichment*

4.3.1.1 *Extracting gene expression data from images*

As outlined in the previous chapter, all gene expression results described here were extracted using 2D expression images from the Allen Brain Atlas (ABA), which represent the degree of probe binding in *in situ hybridisation*(ISH) experiments. I used a Medical Imaging Registration Toolbox (Myronenko and Song, 2010) to register these 2D images to a common reference image such that pixel coordinates mark equivalent locations across images. This enabled me to classify genes into groups based on their regional RNA expression. Where images could not be accurately registered, I relied on the results of ABA tools and a NeuroBlast search, followed by manual validation, to identify genes with patterned expression and assign them to categories.

4.3.1.2 *Calculating the similarity between gene expression in MEC and other cortical regions*

To investigate the relationship between gene expression in MEC and in visual and somatosensory (SS) cortices, I used the reference image for ML3, combined with reference atlas images, to outline the major cell body layers in visual and SS cortex (Fig. 4.1). I chose a smaller area to represent SS layers than visual or MEC layers because there is more uncertainty about the precision of registration there. I used a Python script to extract pixel intensity data and calculate mean pixel intensity for each region.

To identify genes with high expression in MEC relative to visual and SS cortex, I normalised the data for each gene to the sum of the mean across layers for each of the three regions. Genes were selected for the 'MEC-enriched' group if MEC expression was > 0.6 of the total of all three regions. This threshold was chosen by examining the distribution of genes and selecting a value just outside where the bulk of the data lay.

I also used a form of normalisation to assess differences in laminar enrichment across the three regions. Absolute pixel intensity measures may mask relationships in laminar expression between the regions because there is very high variability across different genes. I therefore normalised the data by calculating, for each gene and each region, the relative laminar expression for that region. The absolute pixel intensity of each layer was divided by the sum across the layers. This allowed me to compare, for example, layer II enrichment in MEC (on a scale from 0 to 1) with layer II enrichment in visual cortex (on a scale from 0 to 1).

To calculate the significance of the relationships in enrichment or expression in the three regions I used linear regression analysis and calculated a Pearson's correlation coefficient. I tested for differences in gradient using an ANCOVA test, run using R. I used a significance α of 0.05.

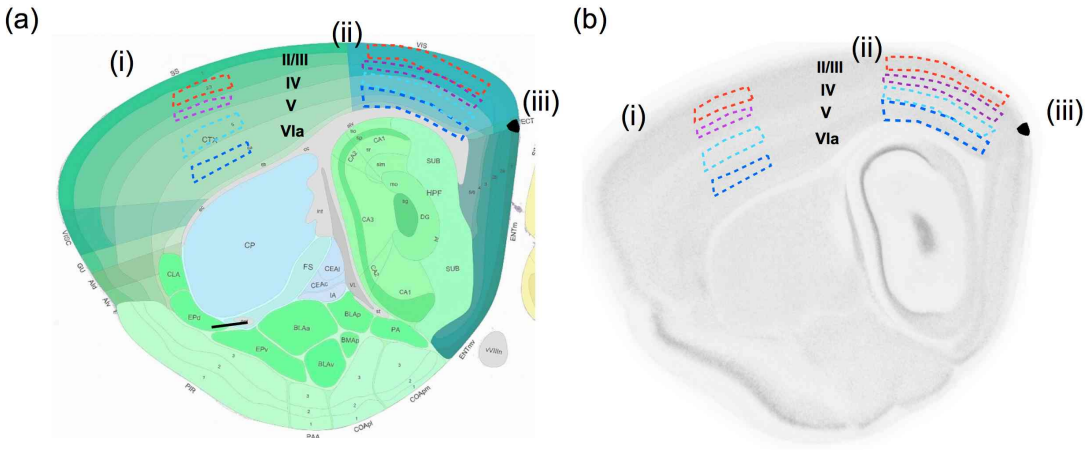


Figure 4.1: *Visual and somatosensory regions used to discover gene expression correlations* (a) Reference atlas image from the Allen Reference Atlas showing the regions selected to measure gene expression in (i) somatosensory and (ii) visual cortex (Vis and SS LII/III, IV, V & VIa), and (iii) in the patches dorsal to MEC layer II. (b) The ML3 reference image used to mark out the layers of (i) SS and (ii) visual cortex, (iii) as well as the large patches. All ISH images were registered to this image.

4.3.1.3 *Procedure for identifying genes with selective expression patterns*

To identify genes with expression patterns that map on to subtle anatomical delineations, either within or beyond the major three cell body layers, required several stages. Initial observations of such genes occurred during my visual assessment for layer enrichment. I used several techniques to identify other genes with similar patterns: correlation score, features similarity and pixel intensity extraction. I used a custom-written Matlab programme to measure the correlation coefficient between images for genes that showed the pattern of interest and all other genes that show expression in MEC. I then visually assessed genes with the highest correlation to build up gene lists of interest. As a second measure I identified the coordinates of the region of interest, for example the large patches dorsal to MEC (Fig. 4.1), and extracted pixel intensity values from this region and the neighbouring region. This enabled me to identify genes with differential expression between the two regions. After visually assessing all possible genes I was able to generate lists of genes with specific expression.

4.3.2 *Analysis of functionally-relevant attributes of enriched genes and their products*

4.3.2.1 *Gene Ontology and Protein pathways*

To examine the enrichment of particular protein attributes in the MEC and its layers, I used the DAVID web facility and its gene ontology database (Huang et al., 2009). This tool extracts ontology terms for each gene then tests if certain ontology terms are overrepresented. Overrepresentation indicates that more genes annotated with a particular ontology are present in a given list than would be expected by chance.

For the majority of the ontology analyses, I grouped individual categories (see Figs. 3.29 and 3.32) into meaningful categories (e.g. all dorso-ventral categories, all layer II-enriched genes). For more specific questions I also performed the analysis on the specific lists. The

nature of the analysis means that the likelihood of obtaining robust enrichment scores is higher for categories with a large number of genes. For the reference set, which is used as a comparison for calculating over-representation, I used all genes in the ABA sagittal database because only these genes had the potential to be included in enrichment lists.

To perform the overrepresentation analysis I used three methods: (1) the DAVID web tool chart report, (2) the DAVID clustering report, and (3) Extraction of gene ontology terms using DAVID, followed by custom analysis. This third analysis could also be used for assessing enrichment of disease genes, custom lists of genes of interest and homology.

1. The DAVID online web tool (Jiao et al., 2012) provides FAT Gene Ontology (GO) terms for the cellular components, biological processes and molecular functions associated with each gene within a particular list. The DAVID tool also extracts molecular pathway information from the Kyoto Encyclopaedia of Genes and Genomes (KEGG) tool (Kanehisa et al., 2004). FAT terms are modified ontology terms that emphasise ontology groupings in the lower branches of the ontology tree, filtering out the broadest terms. There are thousands of FAT terms across the different ontology types, but many are not informative about protein function in neurons. I selected 5 FAT terms of interest for each ontology type and analysed the significance of any differences in representation between each list and the reference set. From the molecular pathways I selected 6 of interest to neuronal function. To calculate enrichment of any type of term, the DAVID tool performs an adjusted Exact Fisher Test (Eqn. 4.2), which is particularly conservative for small lists, followed by a correction for multiple comparisons. I used this calculation to validate results from the custom analysis outlined in (3).

2. One of the major limitations of ontology analysis is the degree of redundancy amongst terms. As such, I made use of an online tool provided by DAVID (Huang et al., 2009) that clusters gene ontology terms across biological processes, cellular components, molecular functions, proteins and upstream sequences, based on their degree of similarity. I have presented the results alongside the results from (1) to provide an unbiased view of the most enriched terms in the layer.

3. DAVID provides an enrichment analysis, but it is limited to overrepresentation. I therefore performed a complementary analysis on the ontology and pathway terms provided by DAVID in which both the probabilities of over- and underrepresentation were calculated. The calculation of significance using the hypergeometric distribution is described below.

The Hypergeometric distribution and Fisher's Exact Test The Exact Fisher test tests the likelihood that a particular set of genes is enriched for a given term, given the shape of the particular hypergeometric distribution. The assignment of gene ontologies can be considered a sampling without replacement problem (such as choosing coloured beads from a bag). This is described by the hypergeometric distribution, the parameters for which are the total number of genes in the reference set (N), the number of genes associated with a given term in the reference set (K), and the number of genes in the selected sample(n). For each annotation, I used the Scipy gauss hypergeometric function to determine the expected value of genes

with the annotation in the sample, given the parameterised distribution. The p value can be calculated by determining the cumulative probability of finding the actual number of genes, given the expected distribution (Eq. 4.2). For a graphical representation of this I used random sampling from the hypergeometric distribution. To perform the final test, I used an R function to implement the Exact Fisher test with 2 tails for under- or over- representation.

$$P = \sum_{i=k}^{n} \frac{\binom{K}{i} \cdot \binom{N-K}{n-i}}{\binom{N}{n}} \tag{4.1}$$

(4.2)

where k *is the number of layer-specific genes of a given GO term,* n *is the total number of layer-enriched genes (sample size),* K *is the total number of genes with a given GO term in the reference set,* N *is the total number of genes in the reference set (http://bioinfo.vanderbilt.edu/gotm)*

Controlling for multiple tests The analysis of gene ontology enrichment involves performing a large number of tests on the same data, and is therefore subject to Type 1 errors, where a correct null hypothesis that there is no enrichment is falsely rejected. A Benjamini and Hochberg (1995) false discovery rate (FDR) adjustment for multiple tests can be applied to the results of such significance tests to minimise the error. Given the hierarchical organisation of gene ontology data, large number of potential categories, and interrelationships between groups, these corrections can be considered overly conservative (DAVID), so I have presented the data based on the uncorrected value, which may already be a conservative estimate for small gene lists, and indicated where the enrichment is significant after the correction has been applied. This correction was calculated using a R function for the disease analysis, homology analysis and genes of interest analysis. For gene ontologies I used the values calculated using the DAVID tool.

Visualisation To visualise over- and under-enrichment, I plotted the $-\log_{10}P$ value calculated using the custom analysis for each group of genes with the appropriate direction. Positive scores therefore demonstrate over-enrichment whereas negative scores demonstrate under-enrichment. I also used the online KEGG pathway tool and DAVID (Huang et al., 2009) for visualising pathway enrichment.

4.3.2.2 *Known gene functions of interest*

Since GO analysis is not specific for neuroscience, I also tested for overrepresentation of known groups of genes, including the post-synaptic density genes, glutamate receptors, ligand-gated receptors, voltage-gated receptors, immediate early genes and G-protein coupled receptors. I obtained lists of receptor genes from IUPHAR database (Harmar et al., 2009), immediate early genes from a paper by Saha et al. (2011) and post-synaptic density genes from Genes 2 Cognition (Croning et al., 2009). I used the methods described previously to calculate overrepresentation and perform a multiple corrections analysis.

4.3.3 *Homology*

To investigate differences in evolutionary conservation between groups of genes, I used the Biomart-Ensembl interface to identify homologs in Humans, *Drosophila* and Yeast and to calculate the similarity between the homologous pairs.

I applied the sampling method described above to establish whether the proportion of genes possessing homologs was within the range of expected values, given the size of the particular list.

4.3.4 *Disease Annotations*

To test whether genes that either show disease-linked mutations or encode proteins that are active in molecular pathways disrupted in disease are enriched in particular regions, I used the KEGG pathway tool (Kanehisa et al., 2004). This tool contains manually drawn pathways of molecules known to be perturbed, either through environment or genes, in certain diseases, as well as molecules that are therapeutic markers or diagnostic markers (Fig. 4.2). These diseases include neurodegenerative diseases, including Alzheimer's disease, Parkinson's disease, Huntington's disease, other nervous system diseases, including epilepsy, as well as cancers, immune system diseases, cardiovascular diseases and metabolic diseases. Annotations for schizophrenia, epilepsy and autism were obtained by searching Gene Cards (Safran et al., 2010).

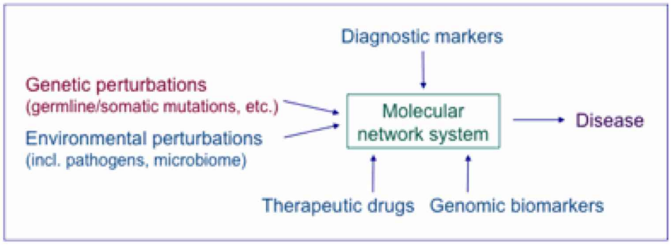


Figure 4.2: *KEGG disease pathway tool* Schematic (from http://www.genome.jp/kegg/disease/) (Kanehisa et al., 2004) showing the types of information used to develop pathway maps for disease.

4.4 RESULTS

4.4.1 *What is the genetic profile of the MEC?*

In the previous chapter I identified a large number of genes that are specifically expressed or enriched within regions of the MEC. In this chapter the aim is to use the data collected to determine the types of genes that are enriched in MEC, how its layers are genetically different and what other patterns of gene expression can be extracted using such a large-scale in situ hybridisation (ISH) resource.

4.4.1.1 *Gene enrichment in MEC follows different patterns to enrichment in visual and somatosensory cortices*

Although the MEC is a cortical region, its layers are organised differently from the main neocortex. I therefore investigated the similarities between MEC and two regions of the neocortex that can be visualised in the same sagittal plane as MEC, the visual and somatosensory (SS) cortices, by examining gene expression in the superficial and deep layers across the three regions of cortex.

Absolute pixel intensity measures suggest that ISH labelling is highly correlated between MEC layers and their equivalent visual and SS cortex layers (Figs. 4.3(A)(a) & 4.4: $r^2 = 0.89 - 0.97, p < 2e^{-16}$). This suggests that variability between different genes is much higher than that across regions for the same gene. Differences in laminar expression between the regions appear to be most pronounced for genes with low expression.

Although a gene may be expressed with similar mean absolute expression in MEC layer II and visual layer II, if expression is uniform across visual cortex, but enriched in layer II of MEC, then this gene may be functionally informative about laminar specificity in MEC based on the protein it encodes. To identify such differences in enrichment between the regions, I selected only genes that show detectable expression in MEC (mean pixel intensity >5), which minimises the contribution of non-ISH noise in the images, and calculated the relative expression across layers in each region.

The relative enrichment in the most superficial layers is very weakly positively correlated between MEC and visual cortex and, to a lesser extent between MEC and SS cortex (Figs. 4.3(A)(b) & 4.4: $r^2 = 0.00 - 0.1$). The slope of the relationship between MEC layer III enrichment and layers II/III of visual or SS cortex enrichment is very shallow. This suggests that genes enriched in MEC layer II may be more likely to be enriched in visual and SS layers II and III than genes enriched in MEC layer III. A more pronounced negative correlation exists between MEC layer V/VI and SS layer V enrichment ($slope = -0.14, r^2 = 0.032, p < 2e^{-16}$), with a slightly weaker slope for MEC and visual cortex ($slope = -0.09, r^2 = 0.02, p < 2e^{-16}$), but the variance explained by each factor is low. In contrast, relative enrichment in MEC layer V/VI and visual and SS layer VIa are weakly positively correlated ($slope = 0.30, 0.19, r^2 = 0.136, 0.054, p < 2e^{-16}$), suggesting that genes enriched in MEC layer V//VI are more likely to show their highest visual and SS cortex expression in layer VIa than layer V. However, these enrichment correlations are extremely low.

To test whether gene enrichment in MEC is more distinct than gene enrichment in visual and SS cortices, I examined the relationship between visual cortex and SS cortex enrichment. The correlations between absolute expression level in layers of visual and SS cortex is also very high with linear regression gradients very close to 1 (Figs. 4.3(B)(a) & 4.4(a): $r^2 = 0.95 - 1, p < 2e^{-16}$). The coefficient of determination is slightly higher for SS cortex than MEC, but results are comparable. However, the correlations between relative laminar visual cortex enrichment and SS enrichment are much stronger than those between visual and MEC enrichment (Fig. 4.3(B)(b) & 4.4(b); ANCOVA $F_{(1,12610)} = 676, p < 2e^{-16}$). Across layers II/III, V and VIa, the correlation between the relative enrichment in that layer of visual cortex accounts for 26-37% of the variance observed for the same layer of SS cortex, compared with other SS cortex layers. It accounts for just 13% of the variance observed for MEC enrichment, and only for the deep layers. Thus, a gene enriched in any layer of visual cortex is more likely to be enriched in the same layer of SS cortex than in an equivalent layer of MEC. MEC layer II enrichment is more strongly correlated with visual layers II/III enrichment than MEC layer III enrichment (Fig. 4.3(B)(b) & 4.4(b); ANCOVA $F_{(1,12610)} = 676, p < 2e^{-16}$). These results suggest that the molecular organisation of MEC layers differs from that of the neocortex, although differences across visual and SS cortex are also present.

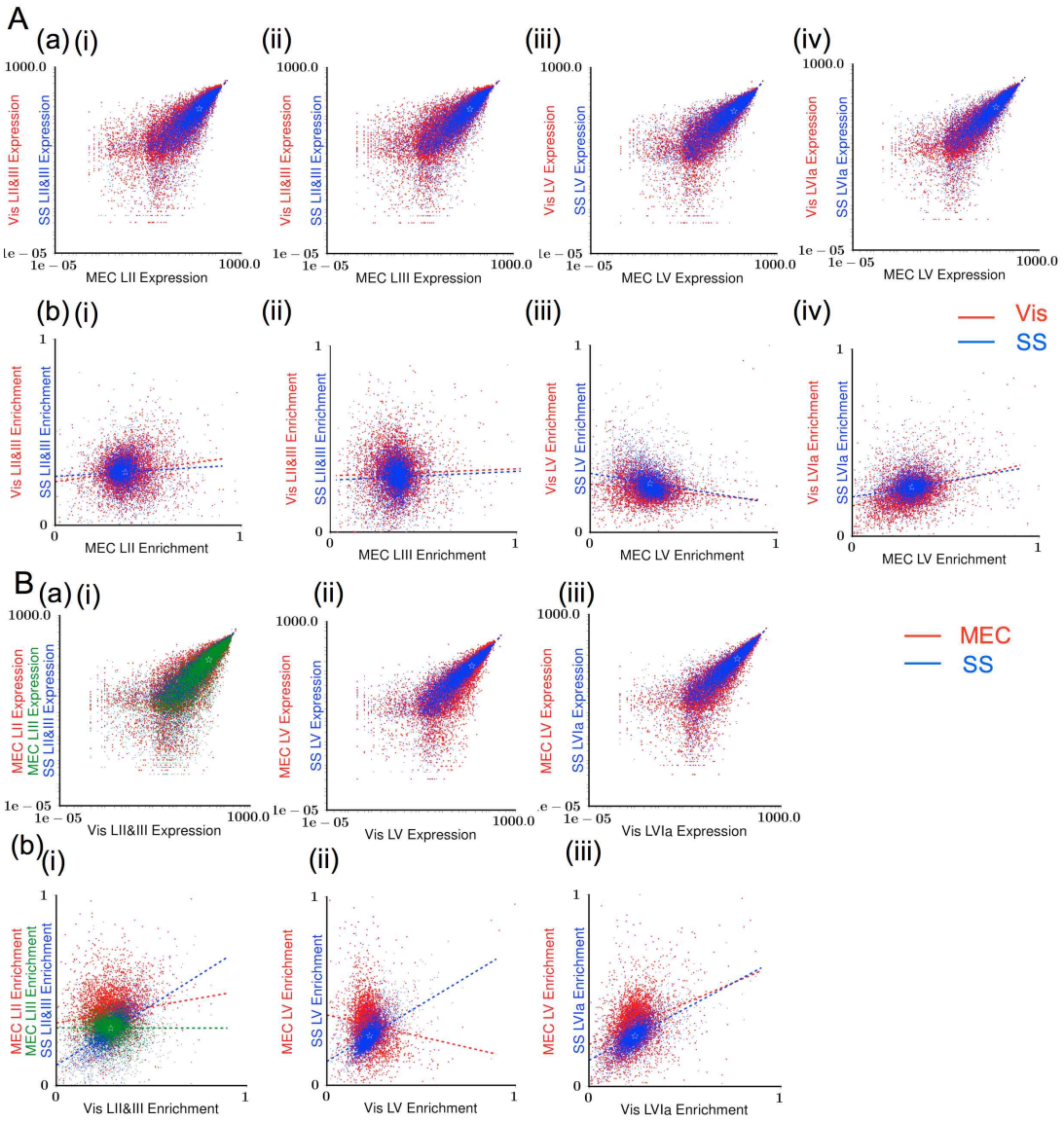


Figure 4.3: *Laminar gene expression is more similar between visual and SS cortices then between MEC and visual or SS cortices* (A) Scatter plots show (a) absolute and (b) relative mean pixel intensity across all pixels in each visual (red) and SS (blue) layer (relative to other layers within the same region), plotted as a function of (a) absolute or (b) relative mean pixel intensity in the respective MEC layer. A log scale has been used for (a) to emphasise differences at low expression levels. Layer comparisons include: (i) MEC LII vs vis/SS LII/III, (ii) MEC LIII vs vis/SS LII/III, (vi) MEC LV/VI vs vis/SS LV, (vii) MEC LV/VI vs vis/SS LVI. For each region pixel intensity was normalised by dividing the layer averages by the sum across the layers. (B) As above, but (a) absolute and (b) relative mean pixel intensity for MEC (red/green) and SS (blue) layers is plotted as a function of (a) absolute or (b) relative mean pixel intensity in visual cortex. The SS layer that directly corresponds with the Vis layer is plotted, whereas for MEC the nearest layer(s) are plotted.

(a) Absolute

X	Y	Grad	r^2	p
MEC II	Vis II/III	1.1	0.9	$<2e^{-16}$
	SS II/III	0.9	0.91	$<2e^{-16}$
MEC III	Vis II/III	1.1	0.89	$<2e^{-16}$
	SS II/III	1	0.91	$<2e^{-16}$
MEC V	Vis V	0.95	0.95	$<2e^{-16}$
	SS V	0.9	0.92	$<2e^{-16}$
	Vis VIa	1	0.97	$<2e^{-16}$
	SS VIa	0.96	0.95	$<2e^{-16}$
Vis II/III	MEC II	0.86	0.9	$<2e^{-16}$
	MEC III	0.9	0.85	$<2e^{-16}$
	SS II/III	0.91	0.89	$<2e^{-16}$
Vis V	MEC V	1	0.95	$<2e^{-16}$
	SS V	0.95	0.97	$<2e^{-16}$
Vis VIa	MEC V	0.96	0.97	$<2e^{-16}$
	SS VIa	0.95	0.97	$<2e^{-16}$

(b) Relative

X	Y	Grad	r^2	p
MEC II	Vis II/III	0.14	0.024	$<2e^{-16}$
	SS II/III	0.062	0.0044	$<2e^{-16}$
MEC III	Vis II/III	0.037	0.001	0.009
	SS II/III	0.047	0.0001	0.015
MEC V	Vis V	-0.092	0.021	$<2e^{-16}$
	SS V	-0.143	0.032	$<2e^{-16}$
	Vis VIa	0.296	0.136	$<2e^{-16}$
	SS VIa	0.187	0.054	$<2e^{-16}$
Vis II/III	MEC II	0.182	0.027	$<2e^{-16}$
	MEC III	0.026	0.0001	0.009
	SS II/III	0.640	0.375	$<2e^{-16}$
Vis V	MEC V	-0.23	0.021	$<2e^{-16}$
	SS V	0.648	0.262	$<2e^{-16}$
Vis VIa	MEC V	0.460	0.136	$<2e^{-16}$
	SS VIa	0.576	0.326	$<2e^{-16}$

(c)

Comparison	$F_{(1,12610)}$	p
MEC II x Vis II/III vs MEC III x Vis II/III	676	$<2e^{-16}$
MEC II x Vis II/III vs SSII/III x Vis II/III	676	$<2e^{-16}$

Figure 4.4: *Gene expression correlations between layers of MEC and visual and SS cortices* (a) Table showing the gradients (Grad), coefficient of determination (r^2) and associated significance p-value (p) for comparisons shown in Fig. 4.3(A,B)(a). (b) As above for Fig. 4.3(A,B)(b). (c) Results of an ANCOVA test to test whether two linear regression gradients are significantly different. All significant correlations are shown in red.

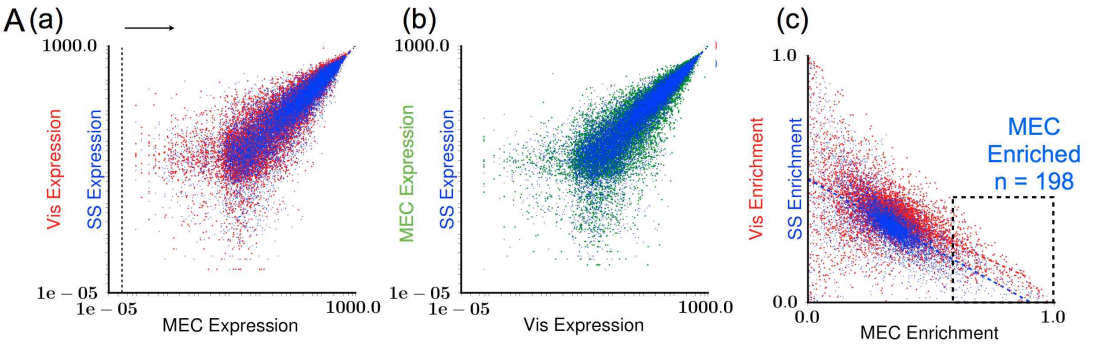
4.4.1.2 *What types of genes show enrichment in the MEC?*

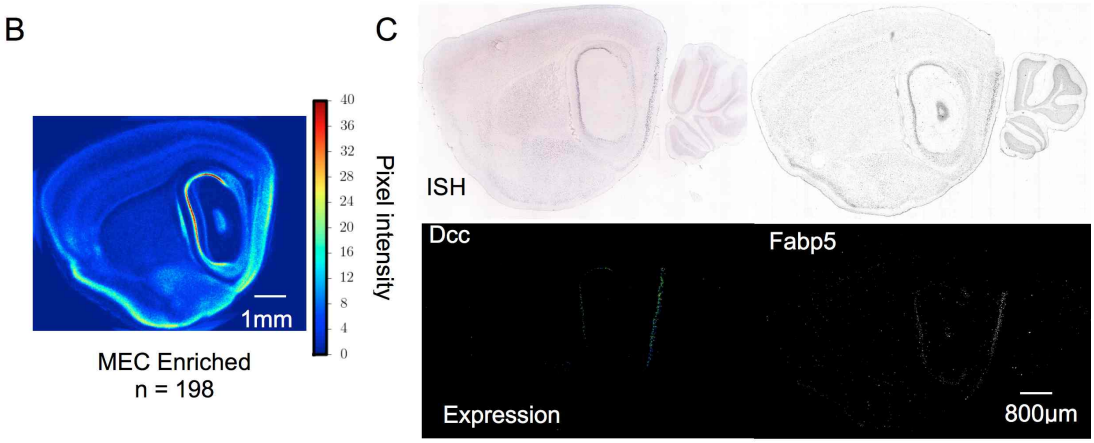
Patterns in laminar gene expression differ between the MEC and neocortex. Are genes that are enriched in MEC compared with neocortex of functional interest, and do their functions map onto specific functions of MEC? I identified genes that are expressed at higher levels in MEC than visual and SS cortices by finding average expression across the layers of each region and calculating the relative expression of each gene across the three regions. The mean absolute pixel intensity measures across the major cell body layers is highly correlated across MEC, visual and SS cortices (Fig. 4.5(A)(a-b): $r^2 = 0.95 - 0.97, p < 2e^{-16}$). However, it is possible to identify genes with very distinct patterns across these regions by calculating relative expression (Fig. 4.5(A)(c)). I used these data to identify genes that are enriched in MEC such that at least 60% of expression is in MEC, with less than 40% shared between the visual and SS cortices (Fig. 4.5(A)(c)).

Genes enriched in MEC typically show highest expression in dorsal MEC layer II and lowest expression in dorsal layer V (Fig. 4.5(B)). On average, their expression level is dorso-ventrally organised within MEC layer II, and stops at the border with the postrhinal cortex. These genes appear to be expressed at moderate levels, on average, in the hippocampus and piriform cortex, and in dorso-lateral MEC, but not in any of the neocortex at this medio-lateral extent. Two examples of MEC-enriched genes that are highly specific to MEC in this sagittal plane, *Dcc* and *Fabp5*, show very low expression across the hippocampus, striatum, and other cortical regions, as shown in (Fig. 4.5(C)).

The enrichment of RNA indicates that certain protein products are more likely to be synthesised in some regions than in others. It is possible to identify the likely attributes and function of the protein by using gene ontologies, which map genes to certain biological processes, cellular locations and molecular functions. I investigated whether genes with particular attributes are enriched in MEC.

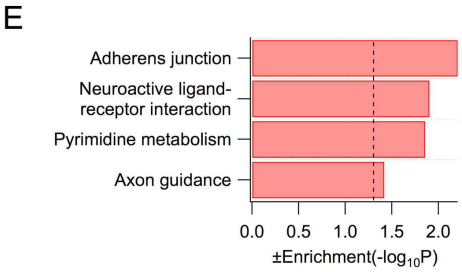
There is little evidence that any functionally-relevant gene attribute is enriched amongst MEC-enriched genes. Genes involved in protein localisation are overrepresented (Fig.4.5(D)), but this is not robust following corrections for multiple testing. A related approach to using gene ontologies is to test whether genes that encode for proteins that function in a specific pathway together are enriched together. Genes involved in the Adherens junction pathway, which maintain cell adhesion, and the neuroactive ligand-receptor interaction pathway, which includes a number of peptides, are slightly overrepresented (Fig.4.5(E)), but post-synaptic genes and voltage- and ligand-gated are present at expected values (Fig.4.5(F)). As such, there is no clear mapping between MEC-enriched genes and protein function.





D

Cluster	Score
PIRSF001238:cyclic amidohydrolases (dihydroorotase, allantoinase)	2.7051
GO:0032507~maintenance of protein location in cell	1.5478
GO:0042923~neuropeptide binding	0.9881
IPR011333:BTB/POZ fold	0.9687
GO:0034705~potassium channel complex	0.9458
GO:0034702~ion channel complex	0.941
GO:0008528~peptide receptor activity, G-protein coupled	0.9302
GO:0043492~ATPase activity, coupled to movement of substances	0.9128
GO:0030097~hemopoiesis	0.8492
GO:0016564~transcription repressor activity	0.8014



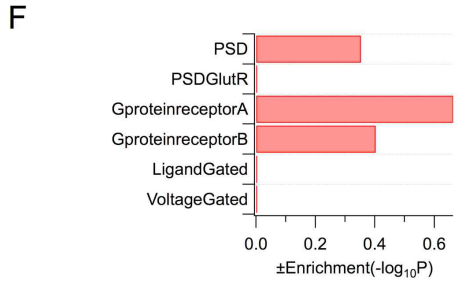


Figure 4.5: *Genes are rarely enriched in MEC compared with visual and somatosensory cortex* (A) (a) Scatter plot showing absolute mean pixel intensity of visual cortex (red) and SS cortex (blue), averaged across layers II/III, IV, V & 6a, as a function of absolute mean pixel intensity in MEC. Genes with mean expression exceeding a pixel intensity of 5 (black vertical line) were subsequently analysed according to normalised expression (see (c)). (b) Scatter plot, as (a), where absolute means for MEC and SS cortex are plotted as a function of the absolute visual cortex mean. (c) Scatter plot, as (a), showing the relative mean pixel intensity of visual and SS cortex (mean across layers) as a function of the MEC relative mean pixel intensity. Pixel intensity was normalised by dividing the visual, SS or MEC layer averages by the overall mean of the three regions. Black box indicates MEC-enriched genes. (B) Image showing average expression pattern for all genes that are enriched in MEC compared with visual and somatosensory cortices (n=198). (C) Images of the ISH images and corresponding expression masks of two example genes enriched in MEC (*Dcc* and *Fabp5*). (D)Table showing the most enriched clustered terms (DAVID) for MEC-enriched genes. (E) Bar plot showing enrichment of the top 4 KEGG pathways for MEC-enriched genes. There are no significant results after correction for multiple tests. (F) Bar plot showing the enrichment of gene groups of functional interest. Dotted line indicates significance at alpha = 0.05 following a 2-tailed Exact Fisher test.

4.4.1.3 *MEC-enriched genes show typical evolutionary conservation*

To investigate whether enriched genes in MEC are more likely than uniform genes to have evolved recently, potentially with the demand for more sophisticated cognitive function, I calculated the extent to which enriched genes possess homologs in *Drosophila*, which are evolutionarily distinct from mice but also have a nervous system, and yeast, a simple organism with no nervous system. I also compared mouse and human homologs to provide insight into the extent to which findings on molecular expression in mice can be extrapolated.

MEC-enriched genes show typical evolutionary conservation of genes included in the Allen Brain Atlas; they possess a similar proportion of homologs across all species measured (Fig. 4.6(a)) and their homologs show similar similarity to the whole reference set (Fig. 4.6(b)).

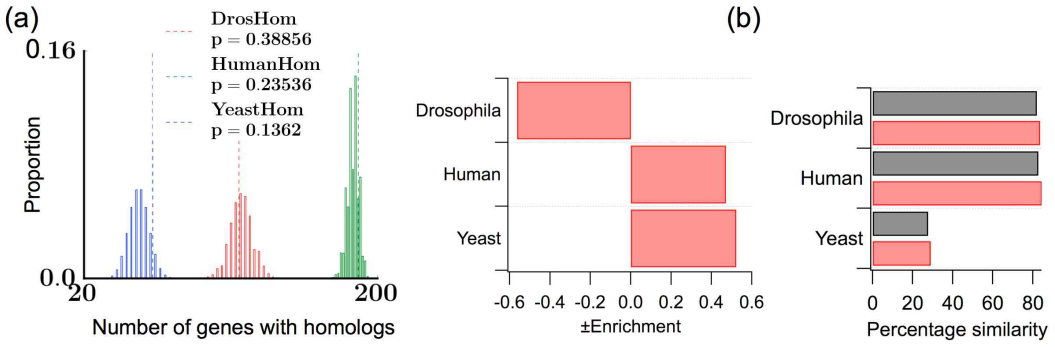


Figure 4.6: *MEC-enriched genes show typical evolutionary conservation* (a) Histograms show the expected distributions of Drosophila (red), Yeast (blue) and Human (green) homologs given a sample the size of the MEC-enriched list. Distributions were calculated using random sampling from a hypergeometric distribution specified by the number of genes in the reference set possessing homologs. Vertical dotted lines show the actual number of genes with homologs in the MEC-enriched list. (inset) Bar chart summarising the enrichment scores for the actual number of genes with each homolog type. (b) Bar chart showing the mean percentage similarity between MEC-enriched (red)*mus musculus* genes and their homologs in each organism, compared with the reference set (black).

4.4.1.4 *Gene expression in MEC and disease*

Abnormal function of the MEC has been associated with a number of diseases. Are enriched genes associated with any of these diseases and could they contribute to the vulnerability of the MEC? Genes that are enriched in MEC compared with visual or SS cortices show a similar likelihood to chance of being involved in the Alzheimer's, Parkinson's or Huntington's disease pathway (Figs.4.7(A)(a) & (B)). They also show average representation of schizophrenia, autism and epilepsy genes (Figs.4.7(A)(b) & (B)), potentially because these diseases affect multiple cortical regions. I investigated the prevalence of disease-associated genes across the brain by plotting the average expression pattern for each disease type gene (Fig.4.7(C)). While genes associated with all diseases show highest average expression in the dorsal CA1 region of the hippocampus, Huntington's disease genes show higher than average expression in piriform cortex and CA3, while HD, schizophrenia and epilepsy genes show high expression across most layers of the neocortex and higher than average expression in the striatum.

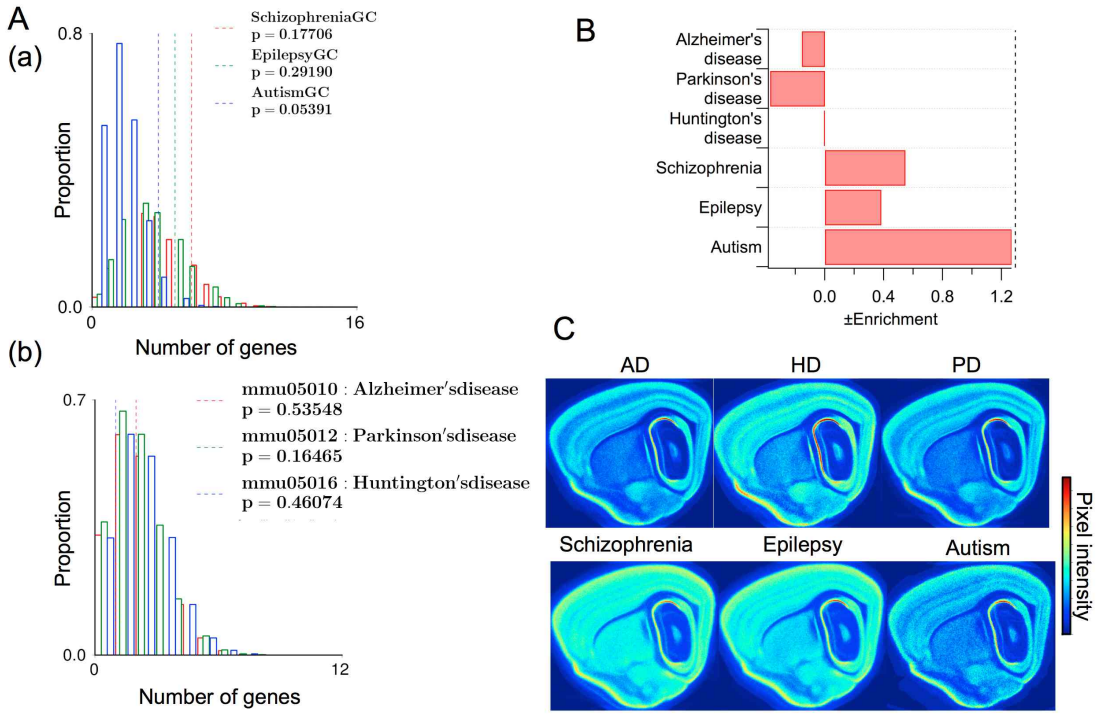


Figure 4.7: *MEC-enriched genes are not enriched for disease annotations* (A) Histograms show the expected distributions of (a) Schizophrenia (red), Epilepsy (green) and autism (grey) genes given a sample the size of the MEC-enriched list. Vertical lines mark actual values. (B) Bar chart showing the enrichment score for each disease annotation. AD, PD and HC annotations were obtained from KEGG pathways (Kanehisa et al., 2004). There are no enriched categories. (C) Average images showing the expression patterns of all genes annotated for a particular disease or pathway. The heat map indicates the level of expression.

4.4.2 *MEC layers show function-related differences in gene expression*

Genes enriched in MEC compared with visual and SS cortices appear to show few overrepresented characteristics. I next investigated whether genes that show specific patterns within MEC, specifically those found be enriched in particular layers in the previous chapter, encode proteins that are associated with particular cellular functions.

While some gene ontology associations are similarly represented across layers, such as ion channel activity molecules, calcium-ion binding genes and synaptic genes, most show considerable differences (Figs. 4.8(A-B)). Dendritic and calcium-binding genes are overrepresented in layer-specific but not uniform or zero-expressing genes, suggesting that these genes strongly underlie laminar differences, whereas protein transport is much more overrepresented in uniform genes.

Only layer V/VI-enriched and uniform genes are enriched for actin-binding and only layer V-VI-enriched for positive regulation of transcription, suggesting that these functions are most important in layer V/VI (Fig. 4.8(B)). Layer V genes are also most likely to be involved in the MAPK signalling pathway (Fig. 4.8(C)) and are most enriched for glutamate receptors in the PSD (Fig. 4.8(D)).

Layer II-enriched genes, which are the most abundant at all levels of specificity, are the most likely to be enriched for particular protein attributes. Only layer II genes are enriched for synaptic plasticity-related functions and learning and memory (Fig. 4.8(B)(b-c)) and are most likely to be involved in the LTP pathway and axon guidance (Fig. 4.8(C)). Genes enriched in layer II are also much more likely to be involved in the citrate cycle and in oxidative phosphorylation, pathways that occur serially in the mitochondria to generate ATP (Fig. 4.8(C)). Over one third of the genes involved in the oxidative phosphorylation pathway are enriched in layer II of MEC. Layer II genes are enriched for PSD and ligand-gated channels but are noticeably underrepresented for voltage-gated ion channels (Fig. 4.8(D)).

Layer III-enriched genes, which are much rarer than LII or LV-enriched genes, show few enriched terms. They are enriched for the intracellular signalling cascade pathway and calcium-ion binding (Fig. 4.8(B-D)), but not robustly enough to be significantly enriched after corrections for multiple tests.

It is also noteworthy than uniformly expressed genes are highly likely to be involved in cellular functions such a protein transport and the PSD, while genes that are not expressed in MEC are highly unlikely to be located in the PSD (Fig. 4.8(B-D)).

A

(a) L II Clusters	Score
GO:0015986~ATP synthesis coupled proton transport	5.292
GO:0030554~adenyl nucleotide binding	4.817
GO:0008137~NADH dehydrogenase (ubiquinone)	3.343
GO:0009201~ribonucleoside triphosphate biosynthetic process	3.259
glycolysis	3.195
GO:0006164~purine nucleotide biosynthetic process	2.935
GO:0008121~ubiquinol-cytochrome-c reductase activity	2.829
GO:0045271~respiratory chain complex I	2.720
GO:0015002~heme-copper terminal oxidase activity	2.572
GO:0007409~axonogenesis	2.567

(b) L III Clusters	Score
GO:0005539~glycosaminoglycan binding	3.007
IPR000909:Phospholipase C, phosphatidylinositol-specific , X	1.973
GO:0005216~ion channel activity	1.483
GO:0034707~chloride channel complex	1.425
IPR018248:EF hand	1.353
calcium-binding region:3	1.334
mmu04914:Progesterone-mediated oocyte maturation	1.227
GO:0004866~endopeptidase inhibitor activity	1.168
cell adhesion	1.146

(c) LV Clusters	Score
GO:0045941~positive regulation of transcription	4.498
GO:0051173~positive regulation of nitrogen compound metabolic process	4.465
GO:0046872~metal ion binding	2.678
PIRSF018476:nuclear factor 1	2.331
IPR013098:Immunoglobulin I-set	1.839
GO:0007409~axonogenesis	1.761
IPR017891:Insulin-like growth factor binding protein, N-terminal	1.569
GO:0005216~ion channel activity	1.522
domain:Ig-like C2-type 3	1.435
domain:PWWP	1.420

B

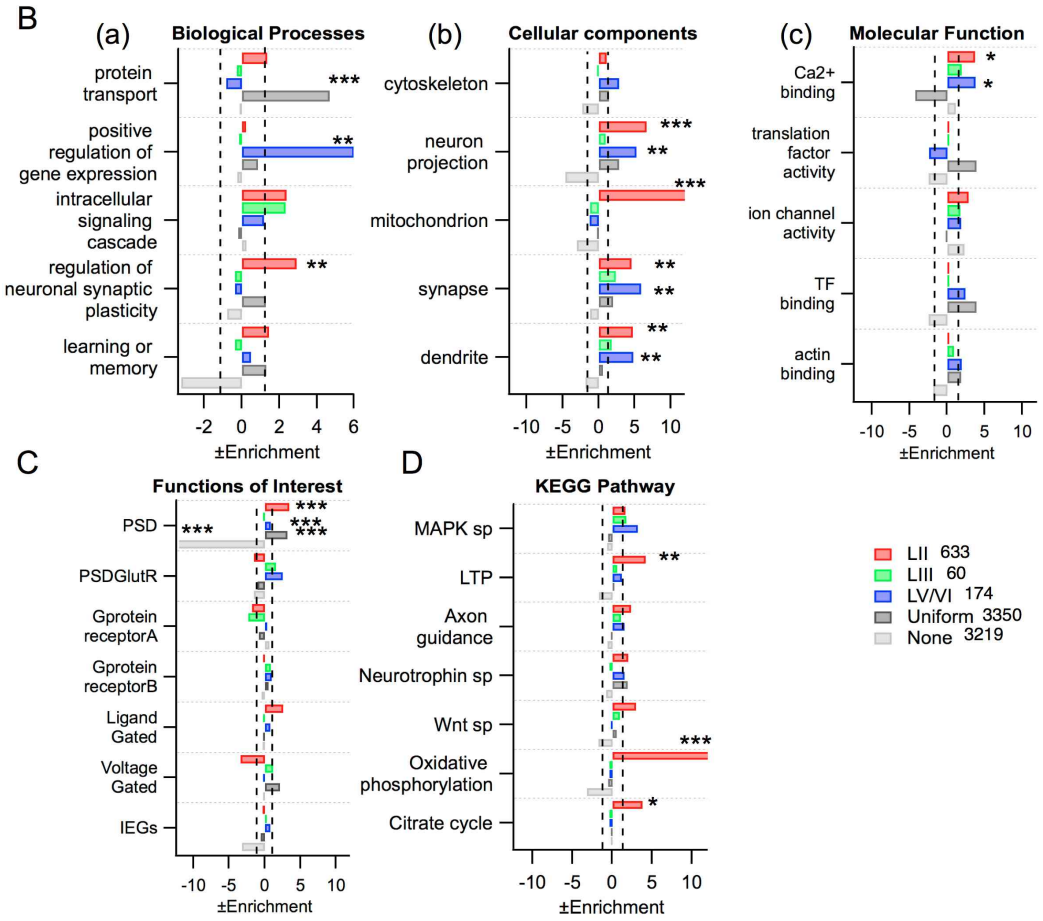


Figure 4.8: *Laminar differences exist in gene attributes* (A) Tables showing the top 10 clustered enriched terms for (a) Layer II-enriched, (b) Layer III-enriched and (c) Layer V/VI enriched genes. Clustered terms are based on grouping of ontology categories and proteins across the classical divisions. (B) Bar charts showing the enrichment of genes annotated with selected ontology terms for (a) *Enriched cellular components*, (b) *Enriched molecular functions*, (c) *Enriched Biological Processes* compared with the reference set. (C) Bar charts showing the enrichment of genes for particular types of proteins of interest. (D) Bar charts showing the enrichment of genes in each list involved in selected molecular pathways. L2enriched genes include genes automatically classified as showing L2only, L2high, L2mid, L232 or L252 expression.L3enriched = L3only, L3high, L3mid, L233, L353. L5enriched = L5only, L5high, L5mid, L255, L355. Dotted lines indicate statistical significance at $p < 0.05$. The number of genes in each category is shown in the legend. The corresponding corrected q value ($p < 0.001 : ***, p < 0.01 : **, p < 0.05 : *$) is indicated for regions in which the overrepresentation is significant after Benjamini and Hochberg (1995) correction.

It is difficult to test for functional enrichment in very small gene lists because the expected number of genes with a particular ontology will be near zero. However, the DAVID cluster tool reveals enriched attributes in small lists by identifying similar terms shared by genes and clustering them together. I used this tool and visual scanning of the lists to identify 3 categories of interest that layer-specific genes appear to fall under. Layer II-specific genes are enriched for neuron projection development molecules (GO cluster 3.88, p = $1.8e^{-5}$, q = 0.0096, n = 8) whereas layer-III specific genes are enriched for cell adhesion molecules (GO cluster 2.13, p = 0.0029, q = 0.146, n = 4). However, both groups of enriched genes include genes from these categories (Fig. 4.9). All three layers show specific expression of ion binding genes, although this group is not significantly enriched in any layer.

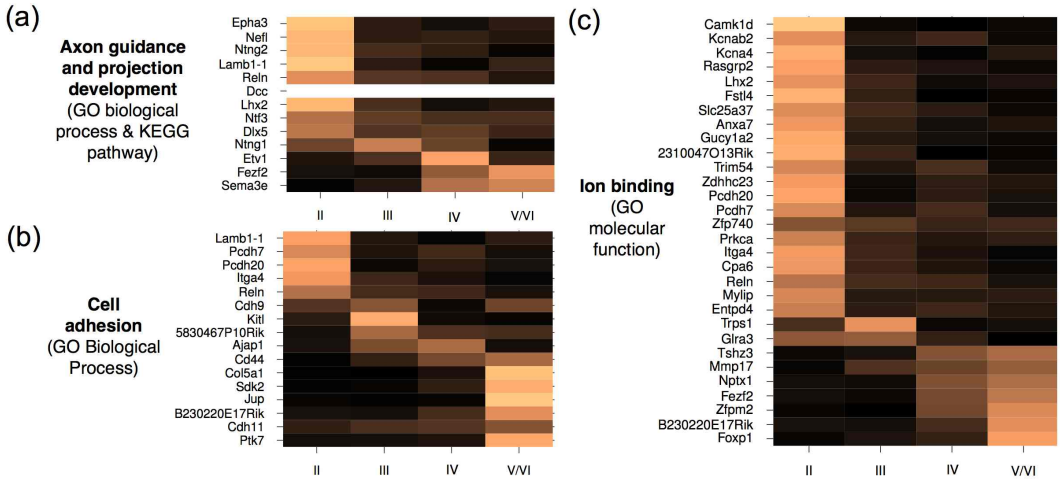


Figure 4.9: *Enriched attributes of layer-specific genes.* Heatmaps are plotted for (a) axon guidance and neuron projection development-related genes, (b) cell adhesion molecules, (c) ion binding molecules across layers II, III, IV and V/VI. Each row (genes) is normalised such that data sums to 1 across layers.

4.4.2.1 *Layer III and V genes are less evolutionarily conserved than layer II genes*

Protein functions are not the only differences between layer-enriched genes. Layer II-enriched and uniform genes are much more evolutionary conserved than would be expected from the reference set, whereas layer III and V genes are actually less likely than chance to have homologs in yeast (Fig. 4.10(A-B)). The similarity between mouse genes and their homologs in humans and Drosophila is very similar across layers but layer II-enriched genes show increased similarity to their yeast homologs, whereas layer V/VI genes appear to show reduced similarity. Genes that show no MEC expression show considerably fewer homologs than expected, possibly because this list includes many genes that have not been well-characterised, which could result in poor probes for ISH and therefore poor labelling (Fig.4.10(B)).

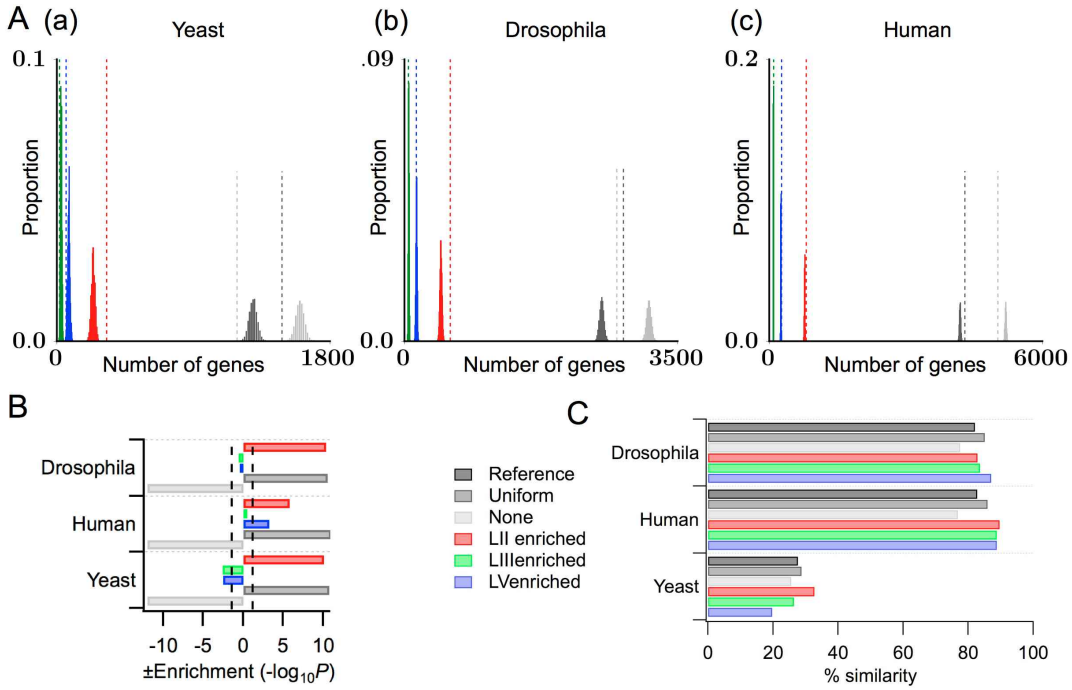


Figure 4.10: *Layer-II enriched and uniform genes show high levels of evolutionary conservation* (A) Histograms showing the distributions of expected number of homologs for (a) yeast, (b) Drosophila and (c) humans for each of the layer-enriched groups. Absolute numbers vary due to differences in the size of each group. Vertical lines mark the actual number of homologs found for genes in that group. (B)(a) Bar chart summarising the enrichment significance of the number of homologs in each organism across different enrichment groups. Dotted lines indicate significance at $p < 0.05$ for over- and under- representation. (b) Bar chart showing the mean percentage similarity between *mus musculus* genes and their homologs in each organism across different enrichment groups.

4.4.2.2 *MEC-enriched genes associated with disease are typically located in layer II*

I showed previously that genes that are enriched in MEC are no more likely than chance to be associated with Alzheimer's disease, Parkinson's disease and Huntington's disease. As shown in Fig. 4.11(A-B), disease annotations are highly enriched in layer II-enriched genes, compared with uniform, layer III or V/VI-enriched genes. In fact, disease annotations for the neurodegenerative diseases are absent amongst layer III-enriched genes. Approximately one third of the genes in the AD pathway are enriched in MEC layer II (Fig. 4.12).

The pattern differs considerably for non-neurodegenerative disorder genes, but this may be because I have made use of genes that are associated with the diseases rather than genes involved in particular molecular pathways. Although layer II genes also show a high proportion of epilepsy and autism-related genes (although the enrichment is much lower), schizophrenia-related genes are more overrepresented amongst layer V/VI-enriched genes, and even layer-III enriched genes (Fig. 4.11 (B)). Images showing the average expression pattern for genes with disease annotations confirm the increased expression in layerII of Alzheimer's disease, Parkinson's disease and Huntington's disease genes whereas for the other diseases expression is more distributed (Fig. 4.11 (C)).

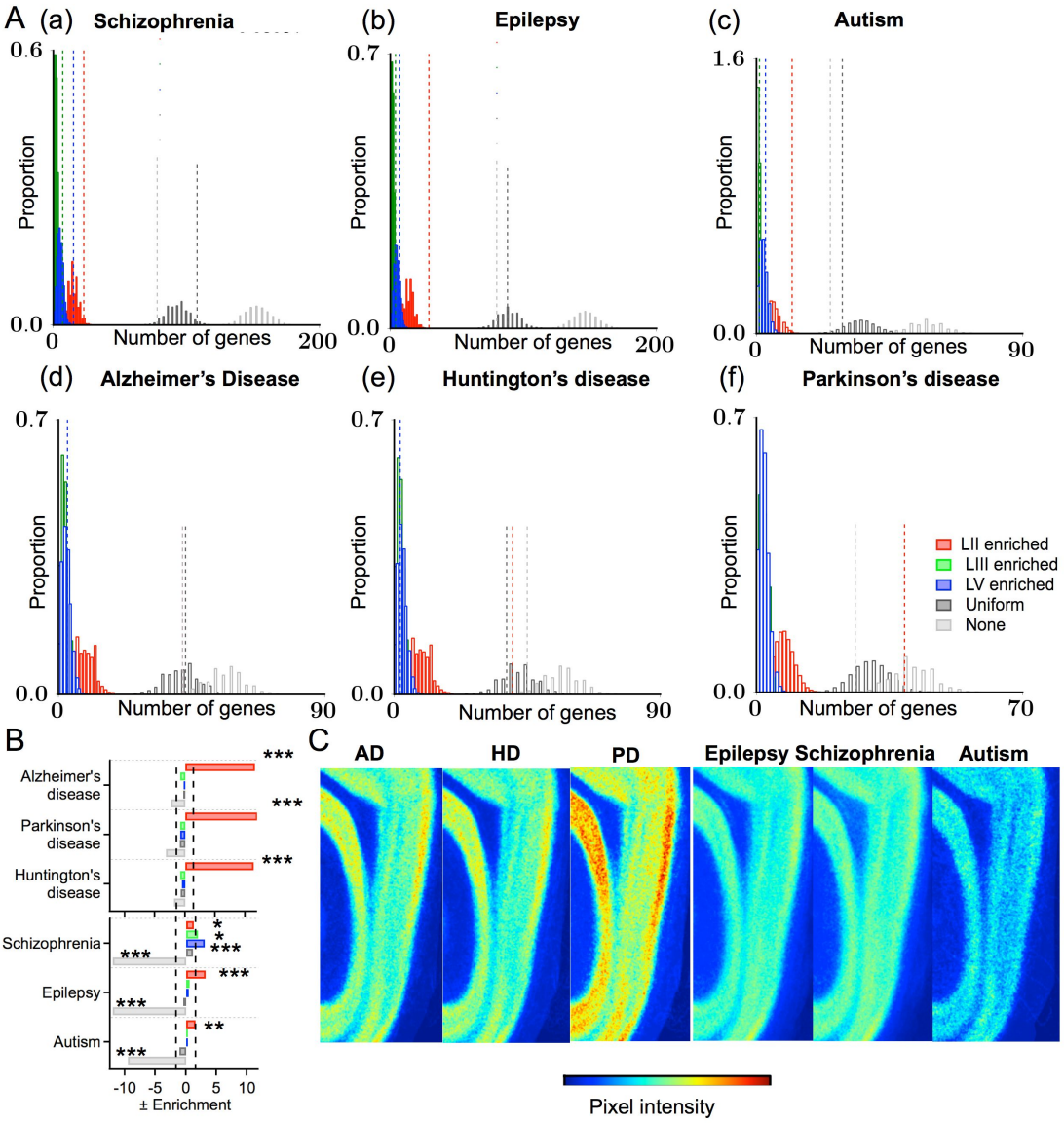


Figure 4.11: *Neurodegenerative disease genes are highly enriched in MEC layer II-enriched genes* (A) Histograms show the expected distribution of genes with a particular disease annotation in each layer after bootstrap analysis, given the total number of enriched genes in the list. The actual number of genes is shown by the vertical dotted line. Data are plotted for (a) schizophrenia, (b) epilepsy, (c) autism, (d) Alzheimer's disease, (e) Huntington's disease and (f) Parkinson's disease. (B) Bar charts summarising the enrichment of disease annotations for each group. AD, PD and HC annotations were obtained from KEGG pathways (Kanehisa et al., 2004). (C) Average images of MEC expression for genes in each group. All data are plotted with the same range of pixel intensity values.

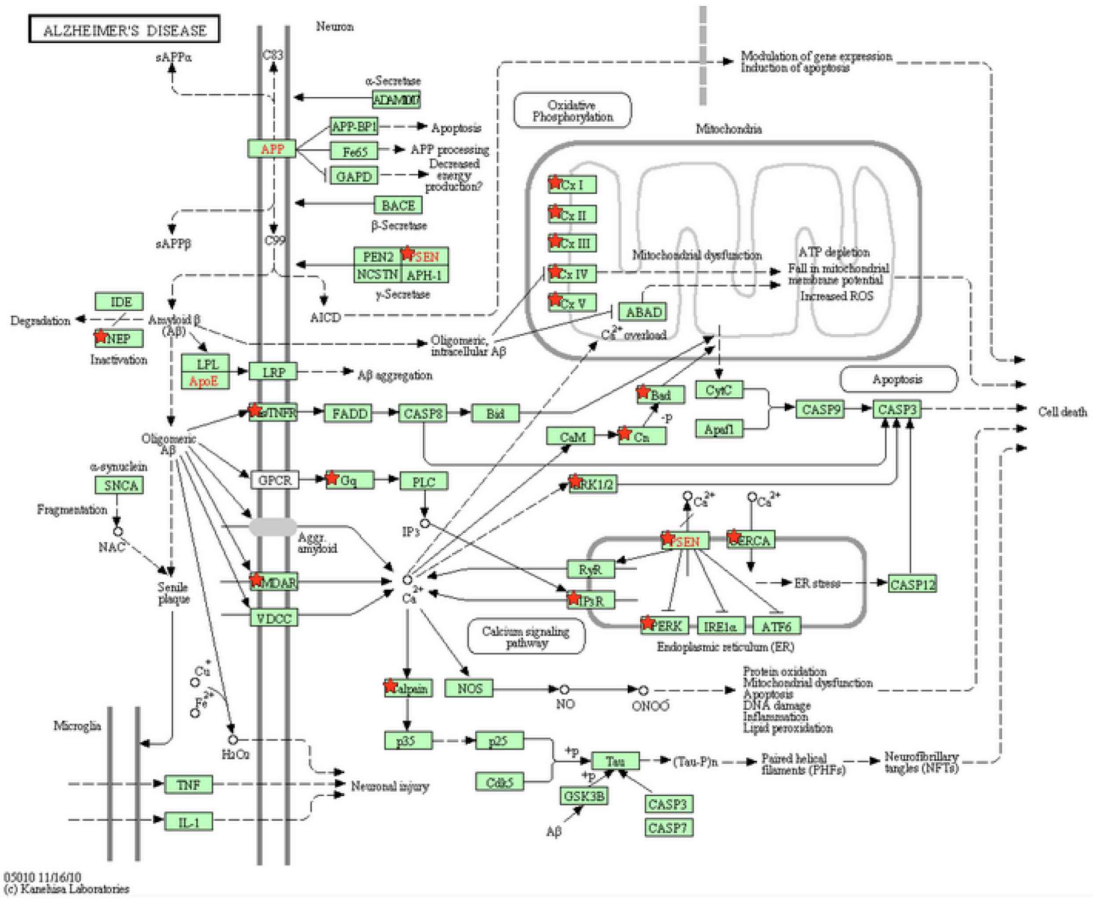


Figure 4.12: *Layer II-enriched genes within the Alzheimer's disease pathway* KEGG pathway image showing the genes within the Alzheimer's disease pathway that are enriched in layer II (red stars). Image obtained from KEGG (Kanehisa et al., 2004) and DAVID, given a list of enriched gene names.

4.4.3 *Does gene expression provide insight into the dorso-ventral organisation of MEC?*

The dorso-ventral organisation of MEC has received considerable attention because cells that are seemingly of the same type respond differently to inputs (Garden et al., 2008) and encode space at different resolutions (Hafting et al., 2005; Stensola et al., 2012). I have previously shown that a large number of genes exhibit dorso-ventral gradients in expression whereas a much smaller proportion show ventro-dorsal gradients (Fig. 3.31). Here, I have investigated the types of genes that are expressed with a gradient.

4.4.3.1 *Dorso-ventral genes are enriched for different functionally-relevant attributes than ventro-dorsal or uniformly expressed genes*

Genes that are expressed with any form of dorso-ventral gradient (DV) in MEC show overrepresentation of a number of different molecular functions, cellular components and biological processes (Fig. 4.13(A)). Mitochondrial components, cytoskeleton-associated genes and neuron projection genes are all strongly enriched amongst DV genes, but are also enriched in genes with no gradient. LTP and Wnt signalling are also common for the two groups, reveal-

ing that certain types of plasticity may occur along a gradient (Fig. 4.13(D)). The LTP pathway is shown in Fig. 4.14; all dorso-ventral expressed genes are highlighted.

Ion channel activity is noticeably enriched amongst DV genes, but not VD or genes without a gradient (Fig. 4.13(B)(b)), as is the intracellular signalling cascade (Fig. 4.13(B)(c)). DV genes are also strongly enriched for MAPK signalling, axon guidance and oxidative phosphorylation pathways (Fig. 4.13(D)). Voltage-gated ion channels and immediate early genes are also weakly overrepresented (Fig. 4.13(C)). Thus, ion channel activity and particular signalling pathways are defining for DV genes.

There are no functional or ontology terms that VD genes uniquely show overrepresentation of, whereas DV genes do not, which may be due to the small number of VD genes. However, it is evident that synaptic genes (and PSD genes, in particular) are enriched amongst all types of genes expressed in MEC (Fig. 4.13(B-C)). This suggests that differences in the expression levels of synaptic genes may drive differences across the regions. At the synapse, the most relevant cellular component for VD genes appears to be synaptic vesicles (Fig. 4.13(A)), which suggests that transmission to other regions with topographical connectivity, such as the layer II-dentate gyrus projections, may be varied by the expression of the genes.

A (a)

DVany Clusters	Score
GO:0008137~NADH dehydrogenase (ubiquinone)	5.772
GO:0007272~ensheathment of neurons	3.540
GO:0008060~ARF GTPase activator activity	3.069
IPR000387:Protein-tyrosine phosphatase	2.396
IPR018316:Tubulin/FtsZ, 2-layer sandwich domain	1.740
IPR005983:Potassium channel, voltage-dependent, beta subunit,	1.686
GO:0030073~insulin secretion	1.515
GO:0005003~ephrin receptor activity	1.493
GO:0030964~NADH dehydrogenase complex	1.452
IPR001431:Peptidase M16, zinc-binding site	1.341

(b)

VD any Clusters	Score
GO:0008021~synaptic vesicle	2.454
domain:C2 2	1.772
GO:0030672~synaptic vesicle membrane	1.710
GO:0035239~tube morphogenesis	1.501
IPR012680:Laminin G, subdomain 2	1.485
GO:0003001~generation of a signal involved in cell-cell signaling	1.273
GO:0030659~cytoplasmic vesicle membrane	1.269
GO:0030814~regulation of cAMP metabolic process	1.197
IPR002048:Calcium-binding EF-hand	1.150
GO:0019933~cAMP-mediated signaling	1.137

B

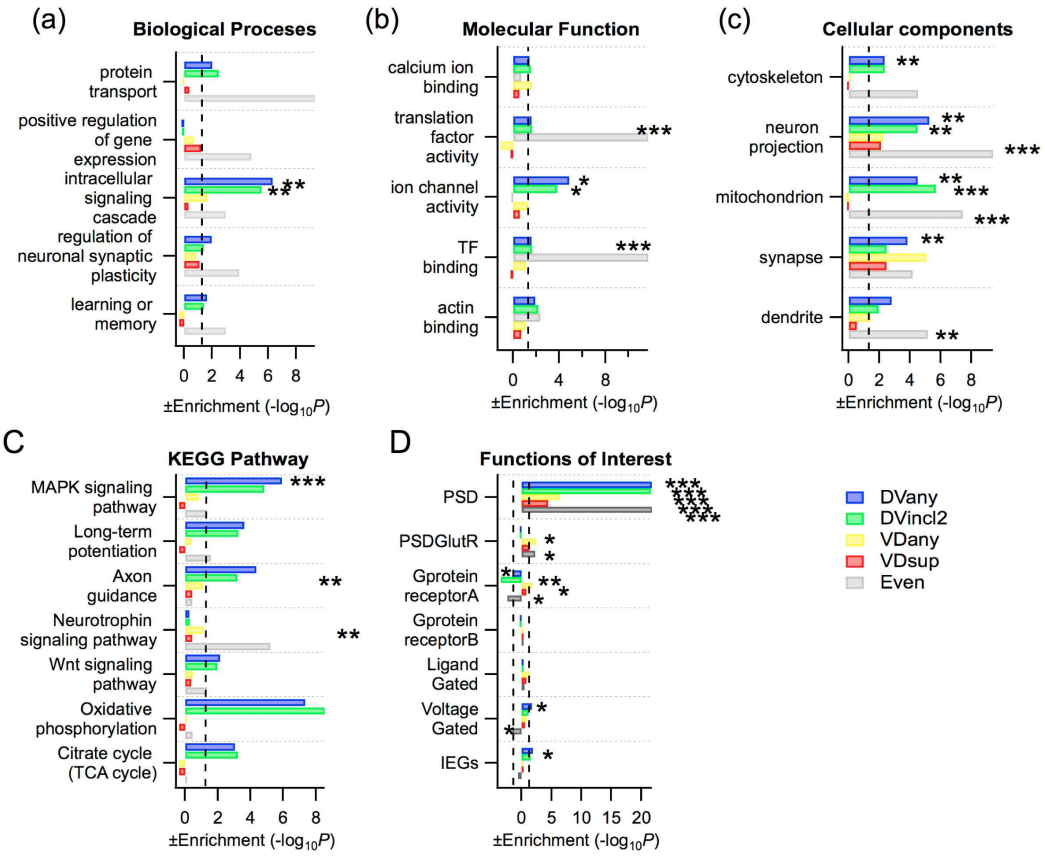


Figure 4.13: *Dorso-ventral genes are enriched for different functionally-relevant attributes than ventro-dorsal or uniformly expressed genes* (A) Tables showing clustered enriched terms (DAVID) for genes with (a) any DV, (b) any VD gradient. Clustered terms are based on grouping of ontology categories and proteins across the classical divisions. (B) Bar charts showing the enrichment ($-\log_{10}$P) of genes annotated with selected ontology terms for (a) *Enriched cellular components*, (b) *Enriched molecular functions*, (c) *Enriched Biological Processes* compared with the reference set. (C) Bar charts showing the enrichment of genes in each list involved in selected molecular pathways. (D) Bar charts showing the enrichment of genes for particular types of proteins of interest. DVany genes (n = 695) include all genes automatically categorised as DVl2, DVl3, DVl5, DVl23, VDDVl2 or DVall. DVincl2 (n = 660) : DVl23, VDDVl2, DVall and DVl2 genes. VDany (n = 80): VDall, VDl5, VDl23, VDl2, VDl3. VDsup (n = 42): VDl23, VDl3, VDl2. Even (n = 3740): Genes that are expressed in MEC but show no DV differences. Dotted lines indicates enrichment at $p < 0.05$. The corresponding q value following a Benjamini correction ($q < 0.001 : ***, q < 0.01 : **, q < 0.05 : *$) is indicated for regions in which the overrepresentation remains significant.

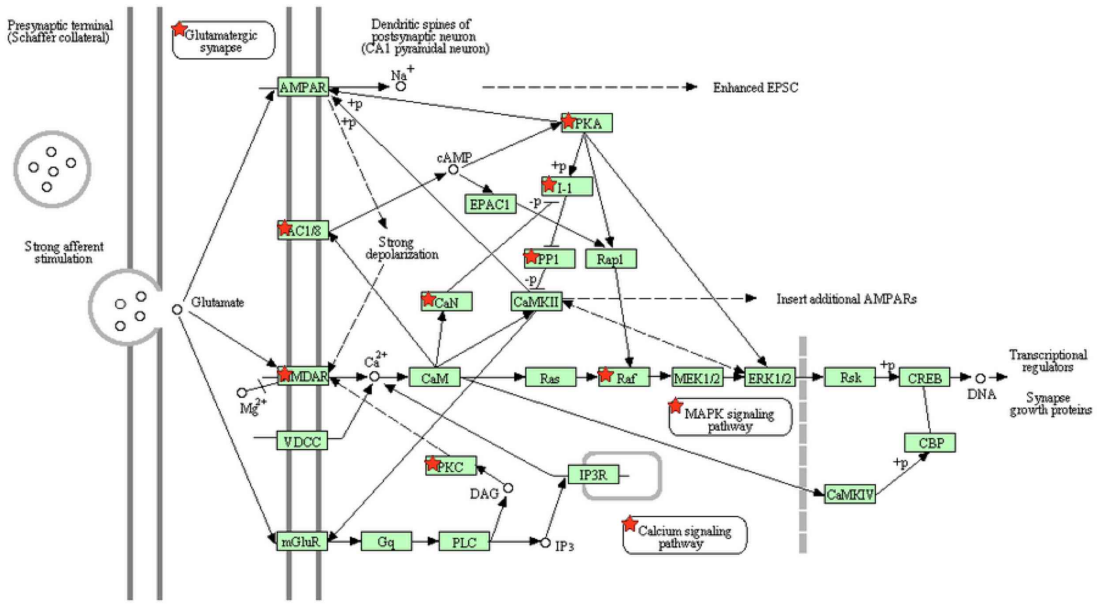


Figure 4.14: *DV-enriched genes within the long-term potentiation pathway* KEGG pathway image showing the genes within the long-term potentiation pathway that are enriched amongst dorso-ventral gradient genes (red stars). Image obtained from KEGG (Kanehisa et al., 2004) and DAVID, given a list of enriched gene names.

Immediate early genes (from (Saha et al., 2011)) are enriched amongst dorso-ventral lists. Since these genes are often indicators of activity and plasticity (Flavell and Greenberg, 2008), I considered the differences in expression pattern. Immediate early genes consistently show a gradient in gene expression (Figs. 4.15(A)(a-b)), that is particularly strong in layer II. The distribution of expression across layers shows higher variability. The gene, *Arc* is known to be necessary for certain forms of NMDA-dependent LTP and LTD (summarised in Bramham et al., 2010). It is regulated by the activity-dependent serum response factor (SRF) transcription factor (Kawashima et al., 2009), which also regulates the expression of other IEGs. The targets of SRF are also enriched amongst dorso-ventral genes (Figs. 4.15(B)), but not ventro-dorsal or genes with no gradient.

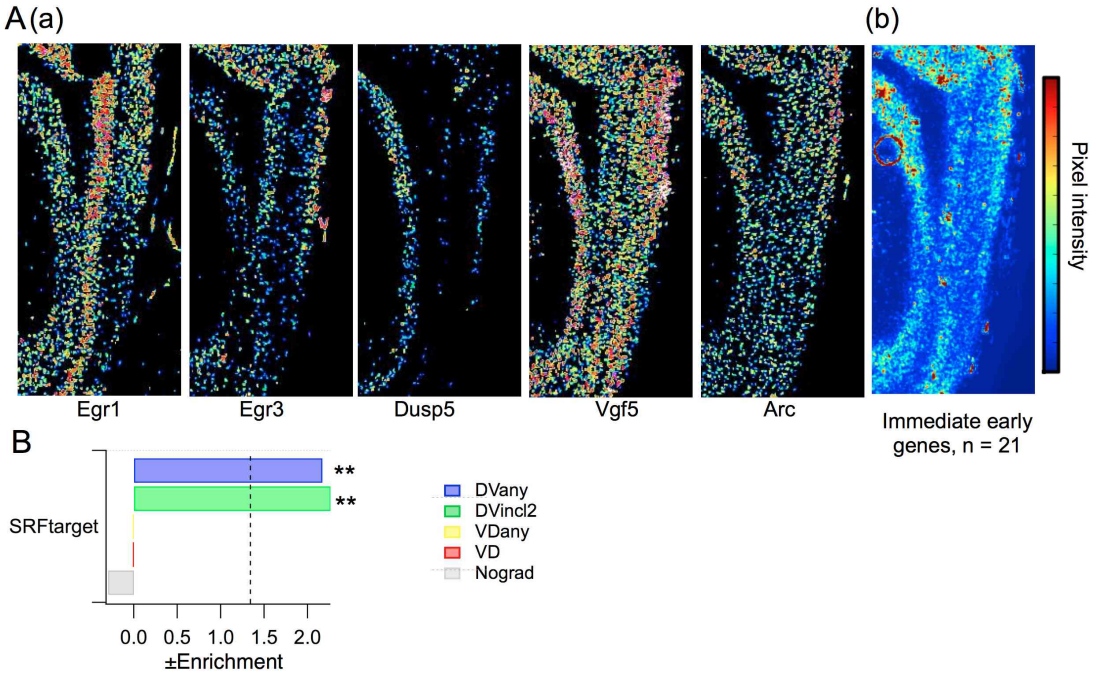


Figure 4.15: *Immediate early genes show consistent dorso-ventral gradients in expression* (A)(a) Registered expression mask images showing gene expression patterns for example immediate early genes (list from (Saha et al., 2011)). While laminar distribution varies, there is a consistent dorso-ventral bias in expression. (b) Averaged expression pattern of all immediate early genes within the MEC region. (B) Bar chart showing the enrichment of SRF-target genes across gradients. Dotted line indicates statistical significance that the number of genes with the annotation is greater than would be expected by chance.

4.4.3.2 *Genes with dorso-ventral gradients are more evolutionarily conserved than ventro-dorsal genes*

I investigated whether genes that are expressed with different types of gradient are more likely to be conserved. Genes with DV gradients and genes with no gradient are much more likely than average, or than VD genes, to possess homologs in yeast or Drosophila, suggesting that they are highly conserved (Fig. 4.16(A-B)). They are also very likely to be conserved in humans. Ventro-dorsal genes that show a gradient in the deep layers, but not necessarily in the superficial layers are less likely than average to show homologs in yeast, suggesting that they are more recently evolved genes. They show increased likelihood of having homologs in humans (Fig. 4.16(A-B)). Genes expressed with a gradient in MEC show higher similarity to their homologs than uniform genes, on average (Fig. 4.16(C)).

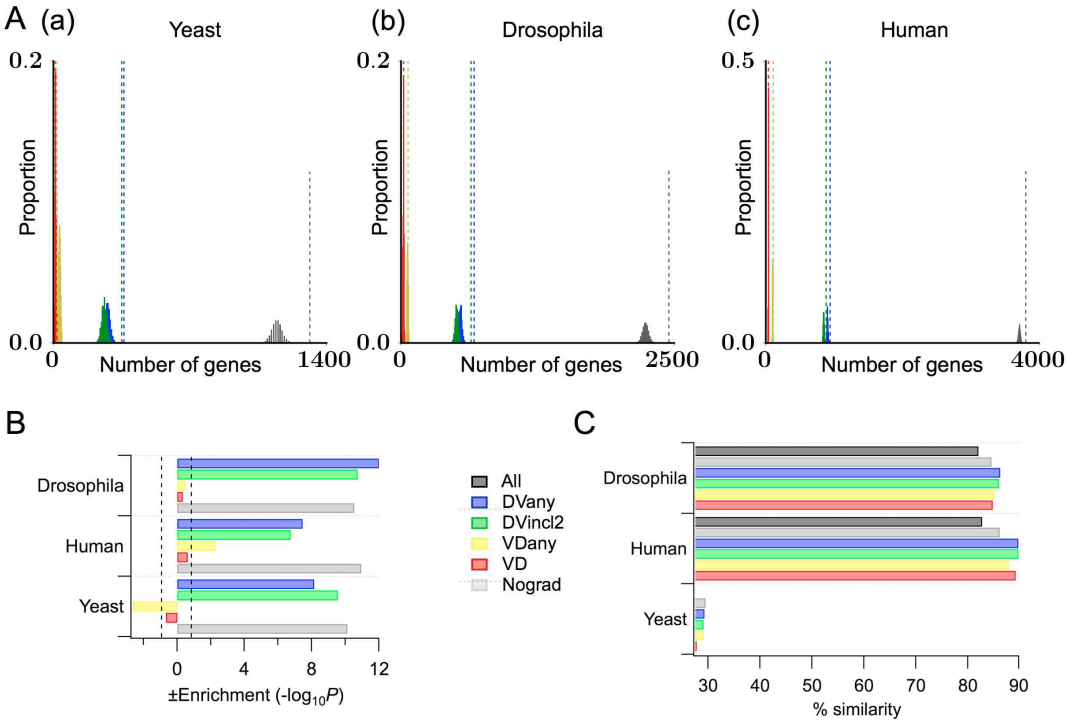


Figure 4.16: *Genes with dorso-ventral gradients are more evolutionarily conserved than ventro-dorsal genes* (A) Histograms showing the bootstrap distributions of expected number of homologs for (a) yeast, (b) Drosophila and (c) humans for each of the layer-enriched groups. Absolute numbers vary due to differences in the size of the group. Vertical lines mark the actual number of homologs found for genes in that group. (B) Bar chart showing the enrichment of homologs in each organism across different DV gradient groups. (C) Bar chart showing the mean percentage similarity between *mus musculus* genes and their homologs in each organism across different enrichment groups. DVany genes include all genes automatically categorised as DVl2, DVl3, DVl5, DVl23, VDDVl2 or DVall. DVincl2: DVl23, VDDVl2, DVall and DVl2 genes. VDany: VDall, VDl5, VDl23, VDl2, VDl3. VDsup: VDl23, VDl3, VDl2. Nograd: Genes that are expressed in MEC but show no DV differences. The number of genes in each category is shown in the legend. Dotted line indicates significance at $p < 0.05$.

4.4.3.3 *Genes involved in neurodegenerative pathways are likely to show dorso-ventral organisation.*

In the previous section I showed that genes involved in neurodegenerative disease pathways are more likely to be enriched in layer II than in any other layer. Genes that are expressed at higher levels in dorsal MEC than ventral MEC are also more likely than chance to be involved in the Alzheimer's disease pathway (Fig. 4.17(A-B)), whereas VD genes show no enrichment. This is also true for Parkinson's disease and Huntington's disease, but zero genes involved in the latter pathway are expressed with a ventro-dorsal gradient.

Is the neurodegenerative disease vulnerability a factor for dorso-ventral and layer II genes, independent of their other features, or are genes enriched in layer II and dorso-ventrally patterned particularly vulnerable? I tested this by comparing gene lists made up of the intersection of these lists with genes not included in at least one list. Layer-II enriched genes are enriched for neurodegenerative disease genes, whether dorso-ventrally organised or not, although enrichment is higher for Alzheimer's disease genes with a gradient (Fig. 4.17(B)(b)). By contrast, only dorso-ventrally organised genes enriched in layer II are enriched for disease genes. The key feature of disease vulnerability therefore appears to be enrichment in layer

II. I wanted to understand how robust layer II vulnerability is, so I tested the enrichment of neurodegenerative disease genes for lists in which the degree of gene expression specificity varies (Fig. 4.17(B)(c)). AD genes are overrepresented in all lists of layerII-enriched genes, but the size of the highly specific list means that the power for detecting enrichment is reduced and there is no significant effect. Nevertheless, in all other lists, including the gene list for highly enriched layer II expression containing just 186 genes, AD genes are overrepresented ($p < 10^{-3.5}$, q (corrected)<0.01).

AD genes have also been shown to be enriched amongst genes with patterned expression that is highest in layer II/III or layer V of somatosensory cortex (Belgard et al., 2011). I compared these lists with a list of all genes showing clear enrichment in layer II of MEC and found that despite the MEC list being smaller, the enrichment is much stronger (Fig. 4.17(B)(d)).

Schizophrenia, epilepsy and autism show different patterns of enrichment in MEC to the neurodegenerative disease genes. Dorso-ventral and ventro-dorsal genes are both highly enriched for schizophrenia genes and epilepsy genes (Fig. 4.17(A-B)), whereas autism-related genes are only weakly enriched in the ventro-dorsal group. When I investigated the breakdown of this enrichment I observed that layer II enrichment is not necessary for the enrichment of schizophrenia or epilepsy genes (Fig. 4.17(B)(b)), but that these genes are even enriched amongst genes specifically expressed in layer II (Fig. 4.17(B)(c)). This enrichment is particularly strong for autism-related genes. Layer II of MEC and layer V of SS cortex demonstrate very similar enrichment of schizophrenia, autism and epilepsy-related genes (Fig. 4.17(B)(d)).

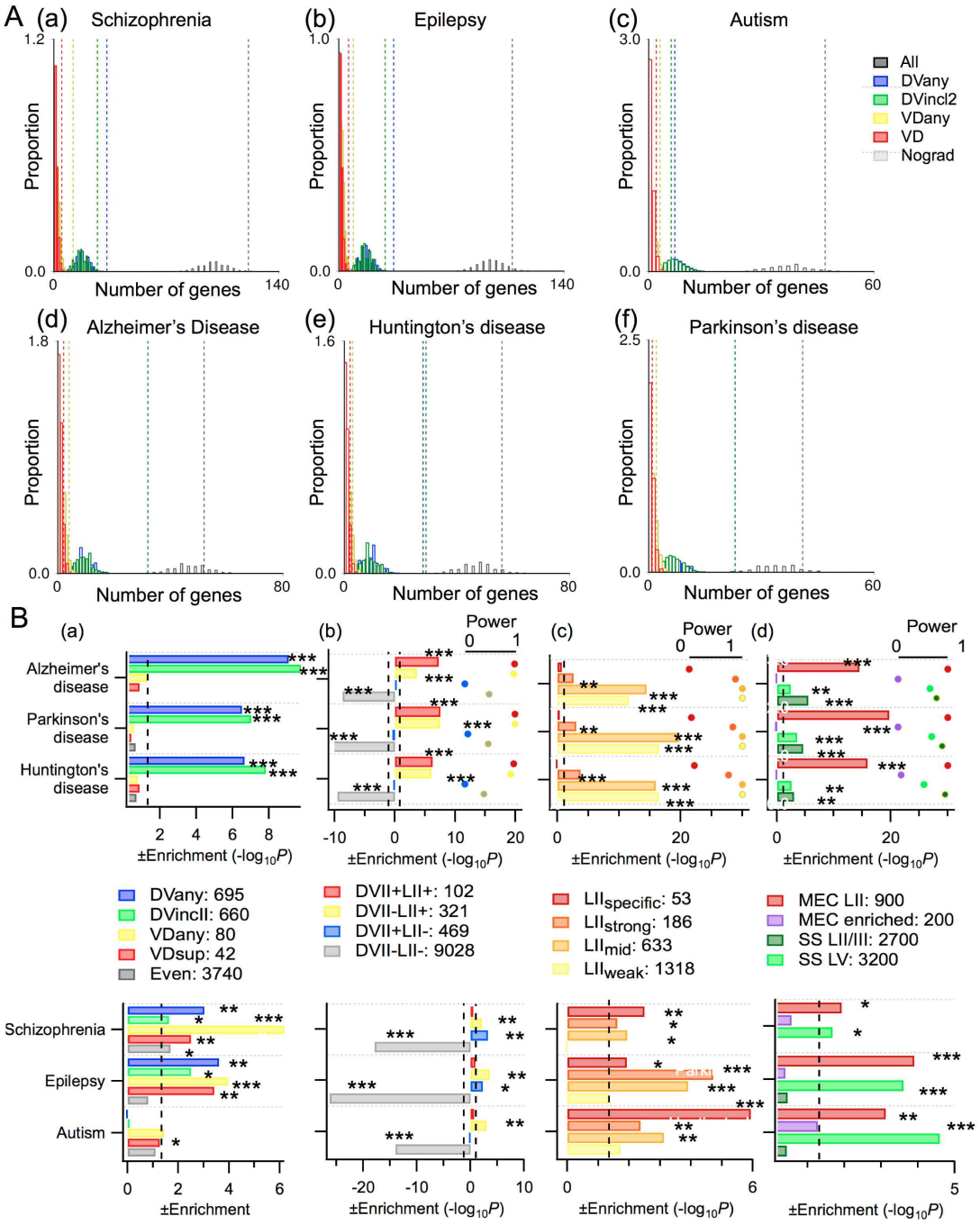


Figure 4.17: *Dorso-ventral genes are enriched in neurodegenerative disease pathways* (A) Histograms show the expected bootstrapped distribution of genes with a particular disease annotation in each gradient group, given the total number of enriched genes in the list. The actual number of genes is shown by the vertical dotted line. (B) Bar charts of enrichment of disease annotations are shown alongside power measurements (dots) for : (a) each DV group, (b) groups of genes that are enriched in layer II and/or with a dorso-ventral gradient compared with genes that do not follow either pattern, (c) genes with different levels of specificity for layer II, (d) genes enriched in layer II compared with somatosensory (SS) layers. SS genes were taken from data provided online by Belgard et al. (2011). AD, PD and HD annotations were obtained from KEGG pathways (Kanehisa et al., 2004). Dotted lines indicate significance at $p < 0.05$:. The corresponding p value after Benjamini-Hochberg correction ($p < 0.001 : ***, p < 0.01 : **, p < 0.05 : *$) is indicated for regions in which the overrepresentation is significant.

4.4.4 *Organisation within layers and functional relevance*

Up to this point, I've considered differences in functionally relevant attributes between reasonably large groups of genes that are enriched in layers consisting of a multitude of cell types. In this section my aim is to examine more closely the expression patterns within layers II and V/VI, and how they interact with dorso-ventral gradients, and to determine whether certain gene functions map onto anatomy.

4.4.4.1 *Expression patterns within layer II and gene attributes*

Cell-type specific gene expression The high resolution of the ABA images that have been used to provide data for this resource may enable identification of sparsely expressed genes that are cell-type specific. In layer II of the MEC several cell types exist, some of which express reelin, and some of which express calbindin (very few express both) (Varga et al., 2010), so I investigated whether either of these cell types could be distinguished.

Calbindin has an interesting expression pattern, showing very high expression in clusters of cells along layer II and zero expression in other regions (Fig. 4.18(A)). By manually searching through visually assessed genes, I identified a number of other genes with this pattern of expression (Fig. 4.18(B)). To attempt to form a molecular profile for this putative cell type, I searched for this pattern across all genes with a form of layer II enrichment.

I classified 'patchy' genes into two groups: genes with very sparse 'patches' expression, and genes with a higher level of expression overall in layer II, but with clearer higher expression in the patchy regions at the layer I border: 'patchy' genes. I did not analyse these genes further, but they may be of interest for understanding what unites and distinguishes cell types in layer II. The averaged image of genes in patches emphasises the consistency of the pattern (Fig. 4.18(C)). The transcription factors Meis2 (Mrg1) and Neurod6 (neuronal differentiation 6) both show a 'patches' pattern. Across the 19 genes that exhibit a strong 'patches' pattern, genes involved in synaptic transmission are overrepresented, along with muscle contraction genes, which on closer inspection is due to the enrichment of calcium signalling pathway genes (Fig. 4.9(D)).

Reelin, in contrast to calbindin, appears to avoid the patches that invade layer I (Fig. 4.18(E)) and is instead expressed strongly throughout the main body of layer II. I also identified a small number of genes exhibiting a similar pattern, which include the potassium channel gene Kcnt1, an interleukin 1 receptor family gene, and an adenosine 1 receptor gene, Adora1 (Fig. 4.18(F)). However, confirming this pattern requires analysis of high resolution raw ISH images because the patchy region cannot be identified in expression images if expression is low. I therefore did not analyse this pattern further.

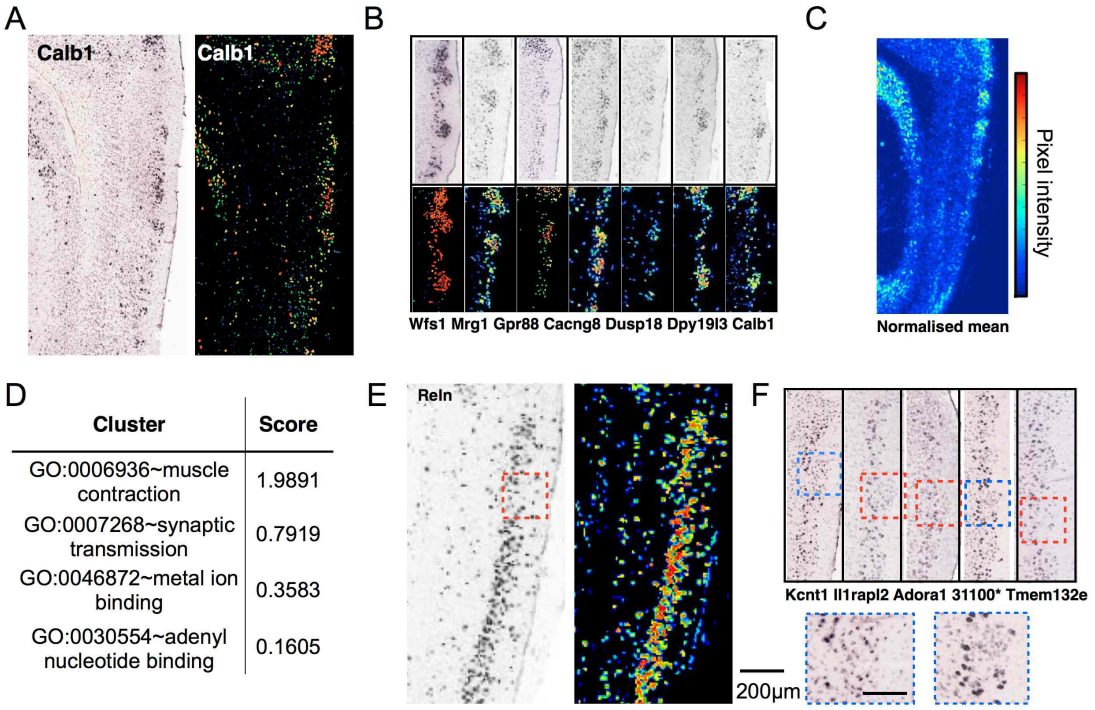


Figure 4.18: *Genes show selective expression along MEC layer II* (A) ISH and expression images showing expression of the gene calbindin. (B) Images showing examples of other genes with a similar patchy expression pattern. (C) Images showing the mean expression pattern for genes with strong expression, only in patches, within layer II. (D) Enriched terms for patchy genes, calculated using the DAVID cluster tool. (E) ISH and expression image showing the gene Reln. (F) Examples of genes with similar expression patterns to Reln in which the patches show low expression. Red boxes indicate patch regions. Blue boxes indicate regions shown at higher magnification (below). Scale bar 100 μm.

Dorsal-ventral gradients within layer II Although DV gradients in grid cell spacing are seen across MEC, cells are most commonly sampled from layer II (Stensola et al., 2012), and *in vitro* electrophysiology work has revealed gradients in the properties of cells in layer II (Garden et al., 2008; Giocomo and Hasselmo, 2009). As such, I examined whether there are specific patterns in the expression of layer II genes along this gradient. Because I wanted to consider precise patterns, I needed a clear understanding of the dorsal border of MEC and whether the large patches dorsal to the MEC proper are sufficiently molecularly distinct to consider them a different population.

Dorsal MEC and the large patches The dorsal border of MEC is not clearly anatomically distinguishable and recent work has posited that the large patches of cells dorsal to MEC form a dorsolateral continuation of the parasubiculum (Boccara et al., 2010; Burgalossi et al., 2011), which is typically considered to occupy the region medial to MEC (Fig. 4.19(A)). Is there evidence in the gene expression data that could distinguish cells in this region from the rest of MEC?

From visual assessment of the ABA data, I identified several genes that showed expression in MEC but not in parasubiculum and several genes with the converse pattern: high expression in the large patches (parasubiculum) but not in the most dorsal region of MEC (i.e. not

continuous expression: Fig. 4.19(B)). Using both a pixel-based search in a manually defined region over the large patches area and MEC and correlational analyses, I identified more candidate, model genes, that show either zero expression in parasubiculum (Fig. 4.19(C)(a)) or high expression in parasubiculum compared with dorsal MEC (Fig. 4.19(C)(b)). The transcription factor Meis2, a 'patches' gene, can also be categorised as a parasubicular/large-patches gene because it is not expressed in the most dorsal MEC region, thereby providing a clear border. Although 'patches' genes occasionally showed this pattern, there were examples in which 'patches' genes were not expressed in the most dorsal MEC or the large patches (e.g. Dusp18).

To explore how these genes are distributed elsewhere in and around the MEC, I plotted the average and variability in expression for the genes (Figs. 4.19(D)). Genes that are enriched in dorsal MEC compared to the large patches are typically expressed in ventral layer V/VI of MEC, in dorsal layer II and in the hippocampus (Fig. 4.19(D)(a)). Genes expressed in the large patches but not in the immediately proximal MEC are typically expressed in neocortex and do appear to be expressed along layer II and V, so the majority of these genes are not large patch-specific (Fig. 4.19(D)(b)). They are, however, typically expressed in more medial parasubiculum (Fig. 4.19(E)). It is likely that other genes with this pattern exist, but this was not the primary aim of this search and it was not systematically performed.

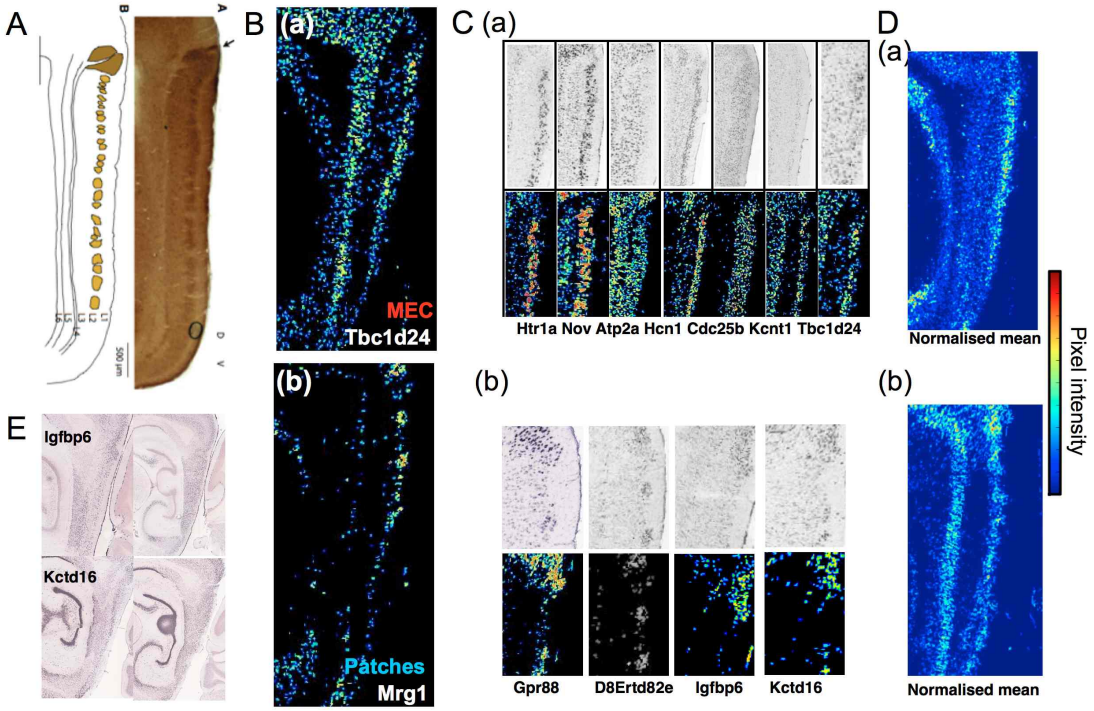


Figure 4.19: *Genes show expression in dorsal MEC layer II that differs from expression in the large patches dorsal to MEC layer II* (A) Schematic and cytochrome oxidase-stained images showing the large patches that make up the putative parasubicular area dorsal to MEC. Adapted from (Burgalossi et al., 2011). (B) ISH and expression images showing initial seed genes for expression patterns that (a) exclude the large patches and (b) include the large patches but not the most dorsal MEC. (C) Images showing examples of other genes with similar patterns that most effectively mark (a) MEC, but not the large patches, and (b) the large patches, but not dorsal MEC. (D) Images showing the mean in expression for (a) MEC, and (b) the large patches. (E) ISH images from more medial sections show parasubicular expression of two of the genes identified as 'large patch' genes.

Could different dorso-ventral gradients in MEC layer II map on to modules? After taking into account changes in gene expression between the large patches and MEC, I identified four different types of pattern in the gradients evident in layer II of MEC that could map on to the modular organisation: (1) gradual dorso-ventral (DV) gradients, (2) steep or abrupt DV gradients, (3) not-dorsal but mid MEC expression, (4) ventro-dorsal gradient (Fig. 4.20(A)).

To distinguish (1) gradual gradients from (2) abrupt gradients throughout the whole dataset, I calculated gradients in pixel expression within the dorsal half of layer II. I selected a threshold gradient of -5 (change in pixel intensity per 1/10th of MEC and thresholded genes based on their mean expression level in MEC layer II. From this list, I identified all the genes manually that showed an abrupt gradient. All genes had to be visually assessed because this 'steep' category is particularly sensitive to registration errors, where ventral MEC is slightly bent out of alignment, and to staining errors and tissue tears. Nevertheless, it was possible to identify a considerable number of genes in this category (Fig. 4.20(B)). Genes in category (2) were defined as those with a strong DV gradient throughout the whole DV extent of MEC, i.e. those without an abrupt gradient.

'Mid' genes (3) were difficult to automatically detect because of slight differences in the coordinates of the dorsal MEC border. For this reason, all such genes were identified during the visual assessment of other lists. These genes can show similar expression patterns to the genes in the previous section that are present in MEC and through absent expression in the large patches, define the MEC-patches border. The defining feature is that they show zero or low expression in the most dorsal region of MEC, and may or may not show large patch expression. This category also shares genes with the 'patchy' category discussed previously. As such, this group is quite heterogenous (Fig. 4.20(B)). Ventro-dorsal genes in layer II were taken from VDall, VDII&III and VDII lists collated in the previous chapter. The differences in the mean expression pattern for the genes in each group is shown in Fig. 4.20(C). Genes involved in synaptic transmission are overrepresented in all groups, although this does not reach significance for 'mid' and 'steep' groups (Fig. 4.13(B), Mid: Cluster enrichment = 1.42, Steep: cluster enrichment = 0.78 (non-significant)).

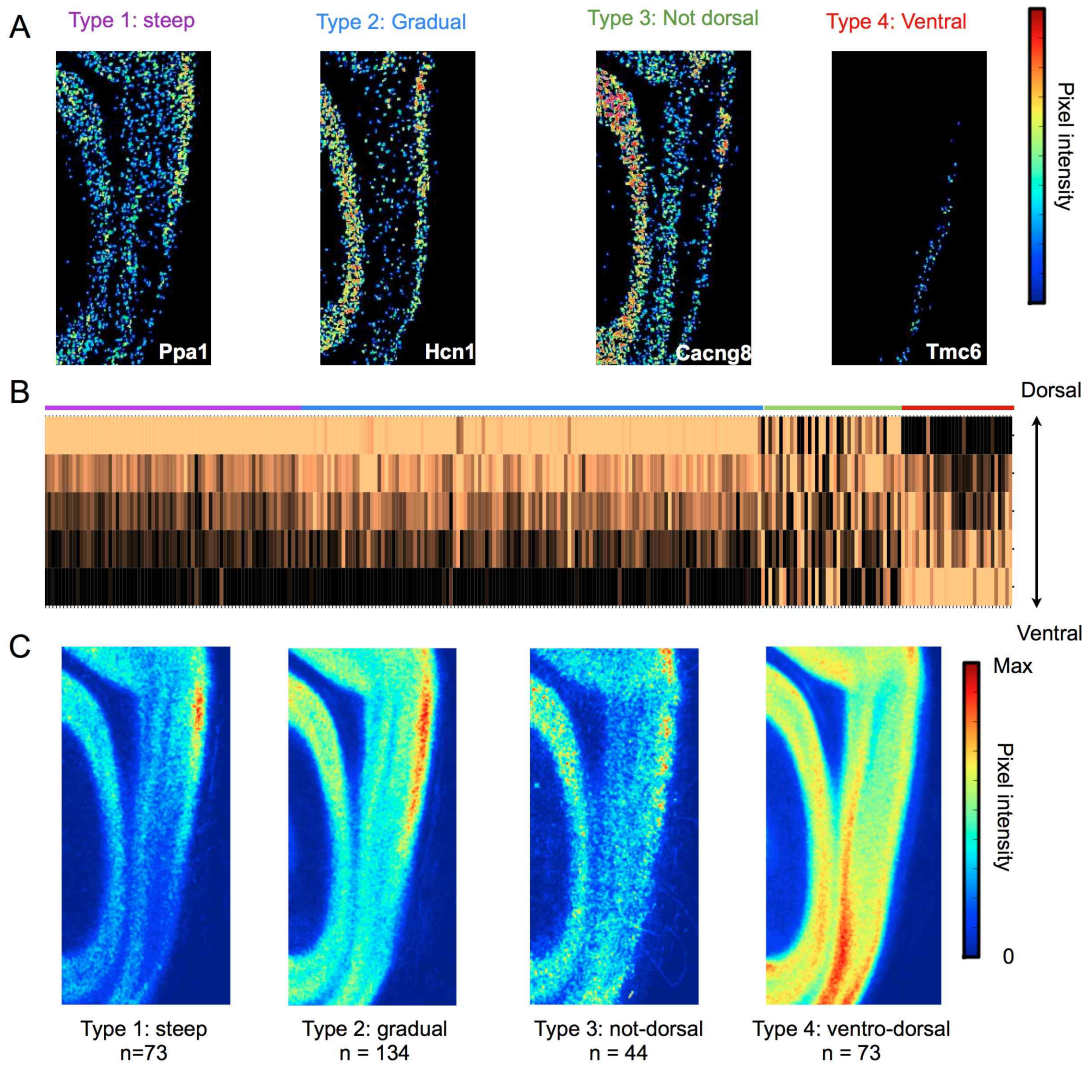


Figure 4.20: *The dorso-ventral organisation of genes in MEC layer II* (A) ISH expression level images of example genes showing different types of gradient in MEC layer II. (B) Matrix heatplot showing the mean RNA labelling level in layers II and III of genes with strong non-uniform expression along the DV axis (columns), for which a valid registered ML3 image is present. Data are plotted throughout the DV extent (rows), resampled from 10 to 5 regions for visualisation purposes. Genes have been ordered according to their expression throughout layer II from steep DV (purple) through gradual DV (blue) through mid (green) to VD (red). Data for each gene have been normalised to the range [0,1]. (C) Images showing average expression for genes classified into the different groups.

4.4.4.2 *Cell adhesion molecules and ion channels that characterise septal CA3 are dorso-ventrally organised in MEC layer II*

It is evident that a substantial number of genes are expressed with a gradient in MEC, but the relationship between dorso-ventral gradients in MEC and septo-temporal gradients discovered in the hippocampus has not been investigated. Thompson et al. (2008) demonstrated that axon guidance molecules, ion channels and transcription factors can all show selective expression throughout the septo-temporal hippocampal gradients, so I examined whether the same genes show gradients in MEC. In all three categories genes expressed at high levels in septal hippocampus typically show their highest expression in dorsal MEC, but tempo-

rally expressed genes do not show clear ventro-dorsal gradients in MEC layer II (Fig. 4.21)), particularly those associated with ion channels.

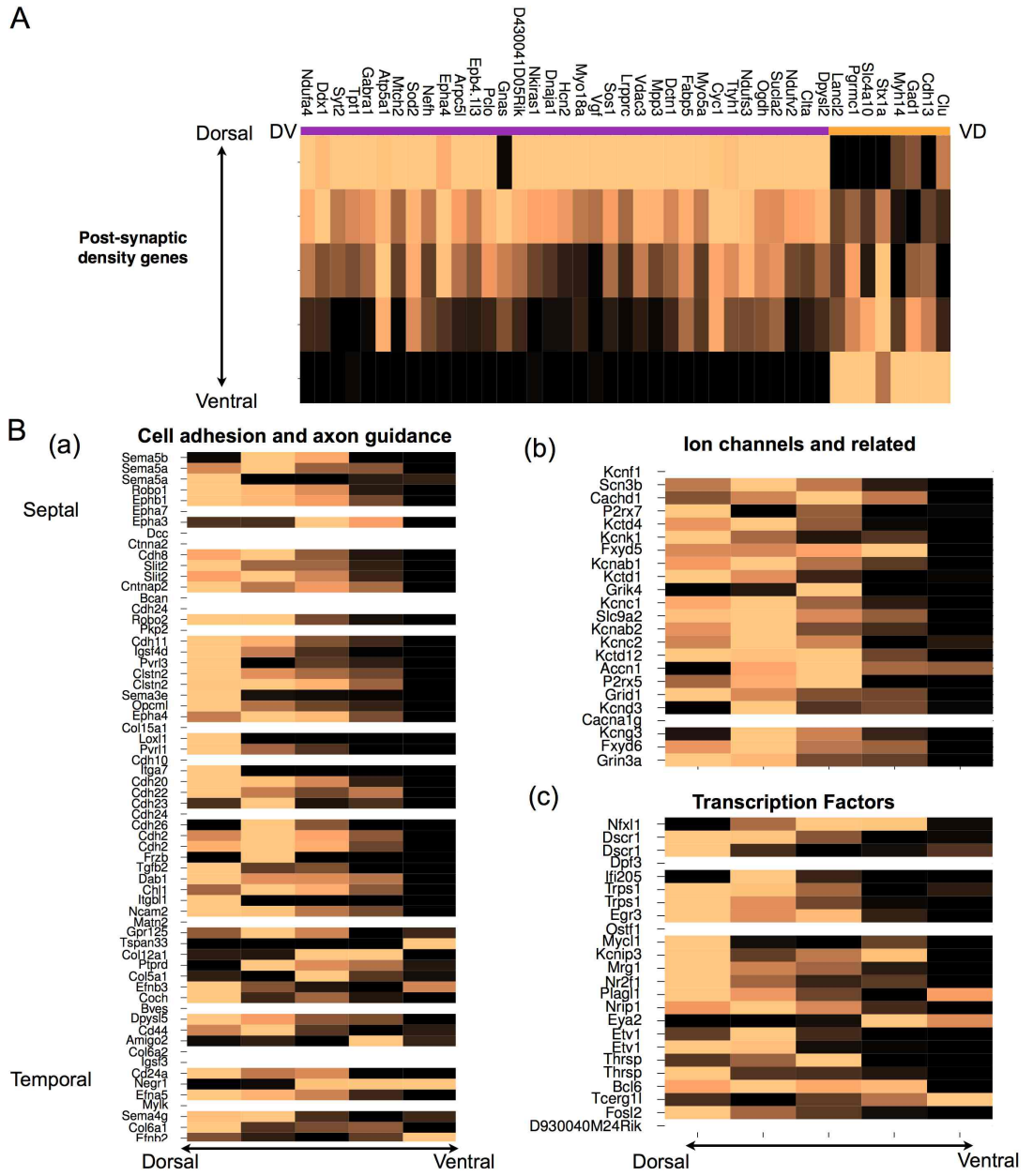


Figure 4.21: *High septal hippocampal expression maps on to high dorsal MEC expression* (A) Heat plot showing the normalised mean pixel expression in the different dorso-ventral regions throughout layer II, resampled from 10 regions to 5 for visualisation, for PSD genes with dorso-ventral or ventro-dorsal expression patterns. All data for each gene have been normalised to [0,1]. (B) Heat plots showing normalised proportion of high-intensity pixels within regions along the DV extent of MEC layer II (columns) for genes (rows) in three different functional categories: (a) cell adhesion and axon guidance molecules, (b) ion channels and related proteins, (c) transcription factors. All data for each gene have been normalised to [0,1]. Regions have been down-sampled from 10 to 5 for visualisation. The same genes were selected and ordered as in Thompson et al. (2008). Genes in the upper region of the heatplots are expressed at higher levels in septal than in temporal hippocampus whereas those in the lower region show the opposing gradient. White rows are present for genes with no high-quality image for ML3.

4.4.4.3 *Mapping out anatomical delineations within layer V & VI and their relation to gene function*

Up until this point layers V and VI have been considered as a whole, but high-resolution ISH data provides a mechanism of identifying genes that are specifically enriched in different layers.

Layer Va Layer Va of the MEC can be distinguished from the main layer V/VI region based on the presence of neurons that are slightly larger than those in layer V and that stain slightly darker with Nissl (Insausti et al., 1997). However, when viewing a section in the sagittal plane, it is difficult to distinguish layer Va cells because it is often only 1-3 cells in diameter (Insausti et al., 1997). I identified a number of genes during visual assessment of NeuroBlast genes that appeared to be specifically expressed or enriched in a narrow band next to the cell body free layer IV (the laminar dissecans), compared with the wider cell body region of layer V (Fig. 4.22(A)). Based on their positioning and distribution I identified these as putatively layer Va.

To find as many genes as possible that show either increased expression in layer Va or specific expression, I used the images possessing this pattern to identify similar genes (Fig. 4.22(B)). To investigate where else in the MEC region these genes are expressed, I plotted the average expression image, and standard deviation in expression image (Fig. 4.22(F)). Genes enriched in the inner band of layer V in dorsal MEC are highly likely to also be expressed in layer II and a narrow band anterior to layer V/VI, which possibly corresponds to a layer VIb, as well as in ventral layer V.

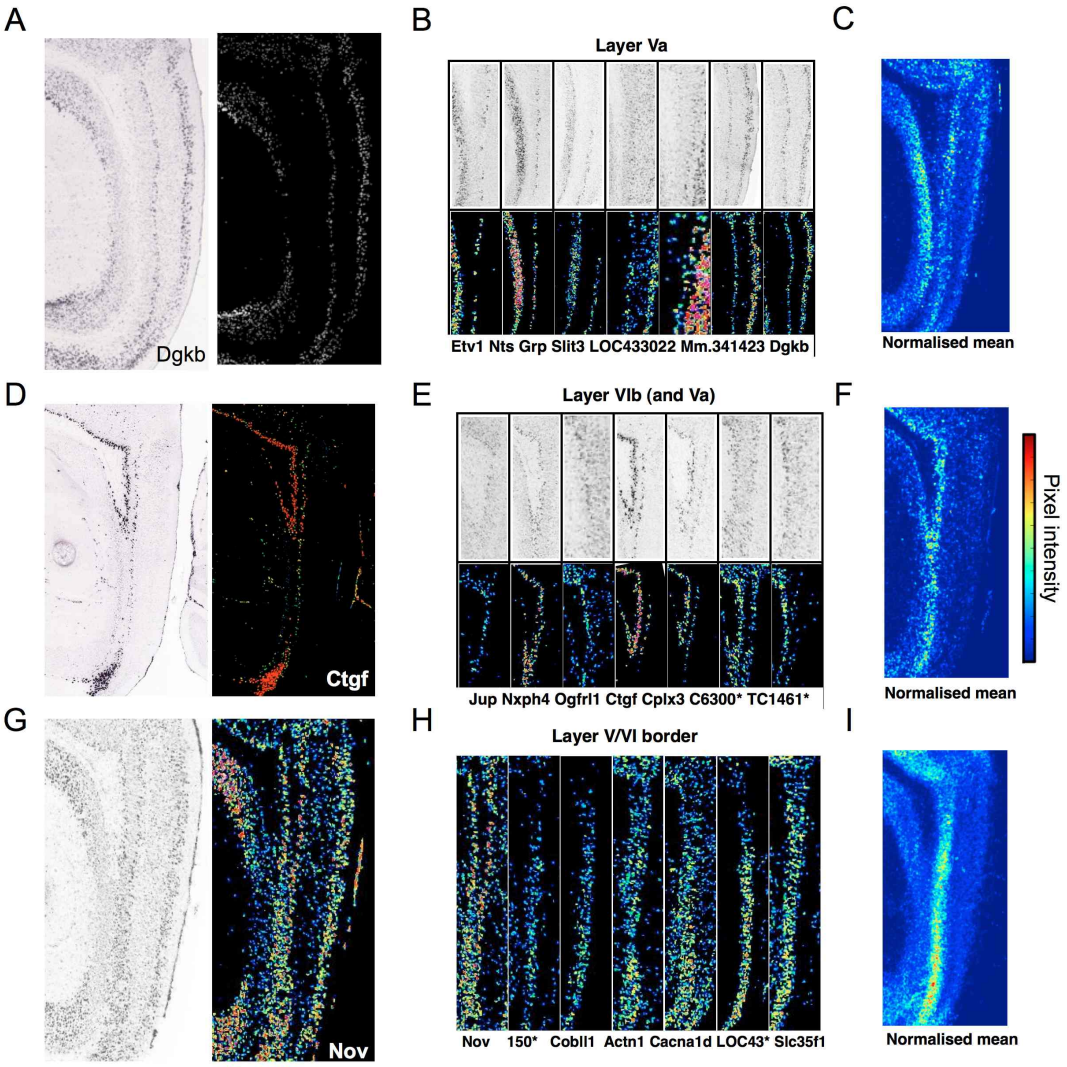
Layer VI Layer VI is not clearly differentiated from layer V except by proximity to the white matter. It contains heterogenously organised cells that intermingle with layer V cells (Insausti et al., 1997) and as such, layer VI is difficult to distinguish from layer V in the sagittal plane. During visual assessment of genes I observed a pattern of high enrichment in a narrow band between layer V and the alveus for a small number of genes, one being Ctgf (Fig. 4.22(D)), a matricellular protein involved in cell adhesion, migration, and proliferation. I used a correlational analysis and feature similarity assessment to identify a number of other genes showing this pattern. Since this pattern is continuous with layer VIb of neocortex (see Ctgf), it possibly demarcates a layer VIb in MEC.

A large proportion of layer VI(b) genes appear to also be expressed in layer Va (Fig. 4.22(E)). Plotting the average expression pattern for a layer VI-enriched gene also reveals that they are often expressed in ventral layer V/VI, particularly in the area corresponding to ventromedial MEC, but expression reduces the more dorsal in the section (Fig. 4.22(F). These genes often show expression in a narrow band corresponding to layer VIb of neocortex and are expressed at low levels in the remainder of MEC, with patchy DV expression in layer II.

Gradients in layer V/VI In the deep layers, unlike the superficial layers, ventro-dorsal gradients amongst genes only expressed in the central layer V/VI region (not the regions corresponding to Va or VI(b)) are more common than dorso-ventral gradients (Fig. 3.31). One possibility is that these gradients are not actually ventro-dorsal gradients but indicate a border between layer V and VI that appears to be variable across animals and planes of sections, and not clear cut, in that genes marking out ventral cells tend to also sparsely invade the

dorsal region. However, the reason it may not be a border is that it is very rare to find genes that are expressed with a dorso-ventral gradient solely in layer V/VI that would correspond to specific layer VI expression. The Neuro-oncological ventral antigen 1 gene (*Nova1*), which encodes an RNA-binding protein, is a rare example of such a gene.

What are the cellular functions of genes enriched across layer V and VI? To gain insights into the protein functions that might be important for distinguishing these regions, I performed an ontology analysis on genes that are enriched in layers Va and VIb. The most enriched low-level groups are shown in Fig. 4.22(J). Plasma membrane genes are overrepresented (GO cc, p = 0.00020, q = 0.016, n = 7), as are neuron projection development genes, but not after correcting for multiple tests (GO cc, p = 0.0023, q = 0.061, n = 5). The biological process terms 'behaviour' and 'cell adhesion' are also enriched, but again they are not robust to corrections for multiple tests (GO bp, p = 0.00037, q = 0.145, n = 7; GO bp, p = 0.0020, q = 0.350, n = 7; 2 genes in both groups).



J

Cluster	Score
domain:CTCK	2.8790
domain:EGF-like 6	1.6493
GO:0031175~neuron projection development	1.4523
domain:Ig-like C2-type 5	1.2732
GO:0007411~axon guidance	1.1321
GO:0016477~cell migration	0.8607
GO:0016023~cytoplasmic membrane-bounded vesicle	0.5365

Functions	Ratio	P
PSD	1.492	0.011
PSDGlutR	-	
GproteinreceptorA	1.679	0.129
GproteinreceptorB	-	
LigandGated	2.674	0.145
VoltageGated	1.622	0.279
IEGs	-	

Figure 4.22: *Genes with selective expression in the MEC deep layer region* (A) ISH and expression image showing initial seed gene for identifying the layer Va pattern (B) Images showing other genes with a layerVa-like pattern that most effectively mark the pattern of interest. (C) Images showing the mean expression for layer Va-enriched genes. (D) ISH and expression image showing initial seed gene for identifying the layer VI(b) pattern (E) Images showing examples of other genes that most effectively mark the layer VI(b) region. (F) Image showing the mean expression for layer VI-enriched genes. (G) ISH and expression image showing an exemplary gene exhibiting a layer V/VI VD gradient or putative layer V/VI border. (H) Images showing examples of other genes that most effectively mark this region difference (I) Images showing the mean expression for layer V/VI gradient or putative layer V/VI border genes. (J) Tables showing enriched terms amongst genes that show patterned expression in layer Va or VI(b) within the deep layers (from DAVID cluster analysis) and the enrichment of particular functional groups of interest.

Other pattern-gene function relationships High-throughput ISH can be a useful tool not only for identifying similarity and enrichment in gene expression over a large number of genes but also for explicitly comparing the expression of genes of interest with others. As an example, I took the calcium-activated potassium channel *Kcnmb4* because it is expressed solely in layer V/VI and its expression behaviour may be related to the calcium-binding proteins that appear to distinguish layers. This type of channel is expressed at very high levels in layer V/VI of MEC but at much lower levels in layer II and III cells, suggesting that it may be specifically important to layer V/VI function. I used a correlational analysis to find genes with correlated expression in the MEC region (Fig. 4.23(A)).

Genes with similar expression patterns to Kcnmb4 include Grik3, the gene encoding a kainate receptor, as well as several calcium-related molecules (Fig. 4.23(A)). Many of these genes also fall into the LV-only category discussed previously, but I also wanted to identify genes with similar expression patterns in the immediate surround of layer V/VI; the subiculum and visual cortex. These genes typically share expression patterns across other brain regions, including the CA1 pyramidal layer (although see BC026657), and across the majority of cortical layers (except Hpcal4, which is absent in layers II/III and V of the visual, parietal and SS cortices).These genes are typically not expressed at high levels in striatum, except for Grik3.

As a second example, I took the interneuron marker parvalbumin (*Pvalb*), which is expressed in a subset of interneurons, including fast-spiking basket cells. Pvalb exhibits an interesting expression pattern that is sparse and spans across layers, but is only expressed at very high levels in dorsal cells (Fig. 4.23(B)(a)). I used feature similarity methods and visual assessment to identify a number of genes demonstrating a similar expression pattern (Fig. 4.23(a)). This similarity can be seen in the average expression pattern across images (Fig. 4.23(B)(b)). An ontology analysis revealed that a larger than expected number of these genes are involved in the regulation of the action potential, including *Pvalb*, *Plp1*, *Cldn11*, *Cd9* and *Serinc5*. Many of the genes are also involved in myelination and are also expressed in cells throughout the white matter tract. This led me to question whether neuron-specific, astrocyte-specific and oligo-dendrocyte specific (as defined by (Cahoy et al., 2008)) genes are differentially expressed within MEC. Oligodendrocytes provide myelination in the CNS. While most astrocytic genes are uniformly expressed and neuron-specific genes show small differences across layers, oligodendrocyte genes are preferentially expressed in layer V/VI and in dorsal MEC (Fig. 4.23(C)). Thus, genes identified as parvalbumin-like may in fact be oligodendrocyte markers.

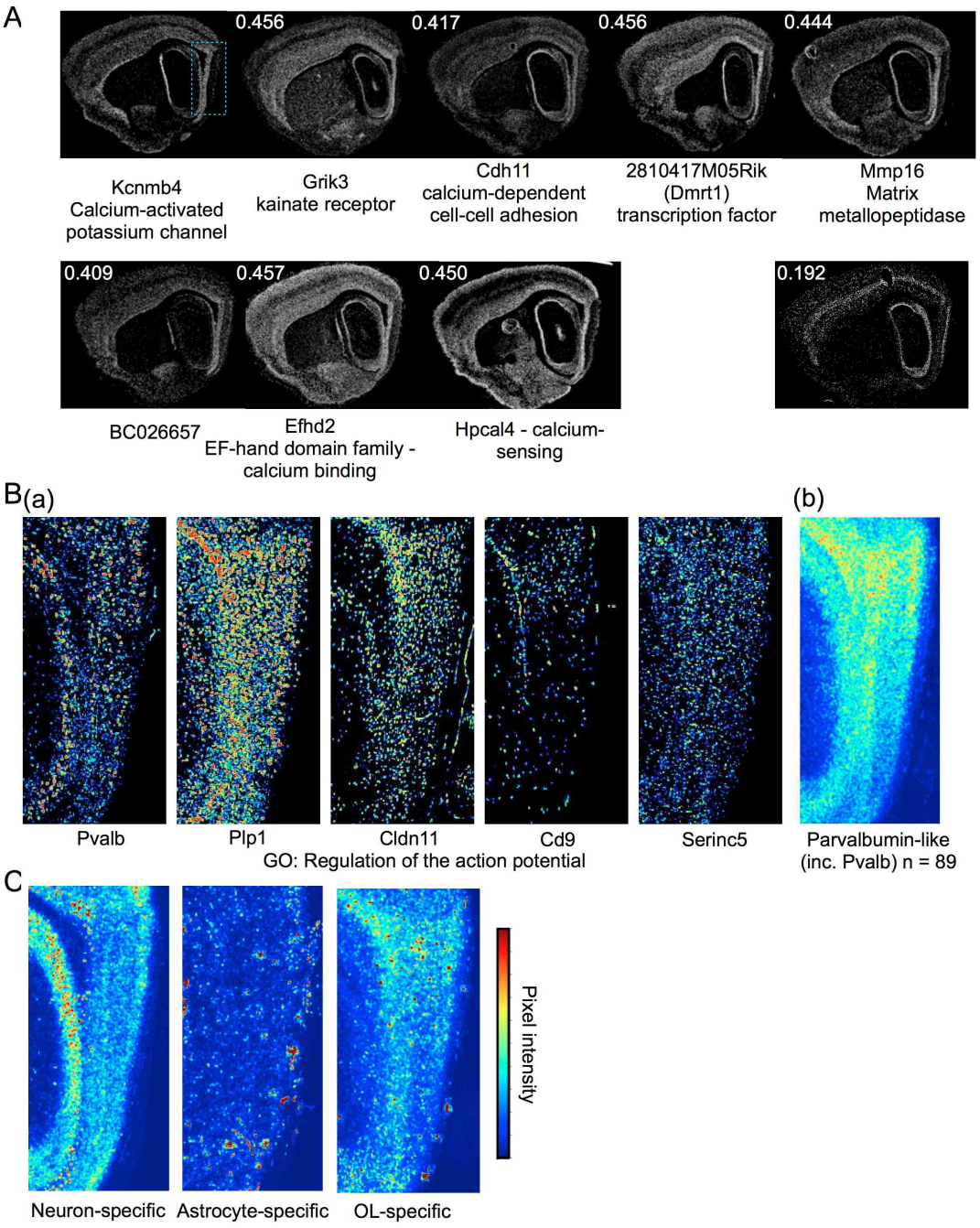


Figure 4.23: *Pattern- gene function relationships* (A) Registered expression mask images showing expression patterns across the whole section for genes with high correlation (indicated) to the ion channel *Kcnmb4* within the MEC region. For a comparison, an image of a layer II-enriched gene with low correlation is also shown (right). (B) (a) Registered expression mask images of the MEC region showing genes with similar patterned expression to the interneuron marker *Pvalb*, that are involved in regulation of the action potential (enriched GO term) (b) Averaged expression pattern of all genes with a similar expression pattern to *Pvalb* (C) Average expression images for neuron-specific, astrocyte-specific and oligodendrocyte-specific genes. Genes taken from Cahoy et al. (2008).

A revised atlas of MEC based on gene expression patterns Through observation and analysis of gene expression patterns it is possible to identify genes that are not only expressed specifically in MEC, but that are likely to define boundaries between layers within MEC, between medial

EC, ventromedial EC and lateral EC and between the MEC and other brain regions, such as parasubiculum. These patterns have enabled me to develop a revised atlas, appropriate for use with sagittal sections in the ML3 plane, that may more accurately reflect laminar differences than current reference atlases (Fig. 4.24).

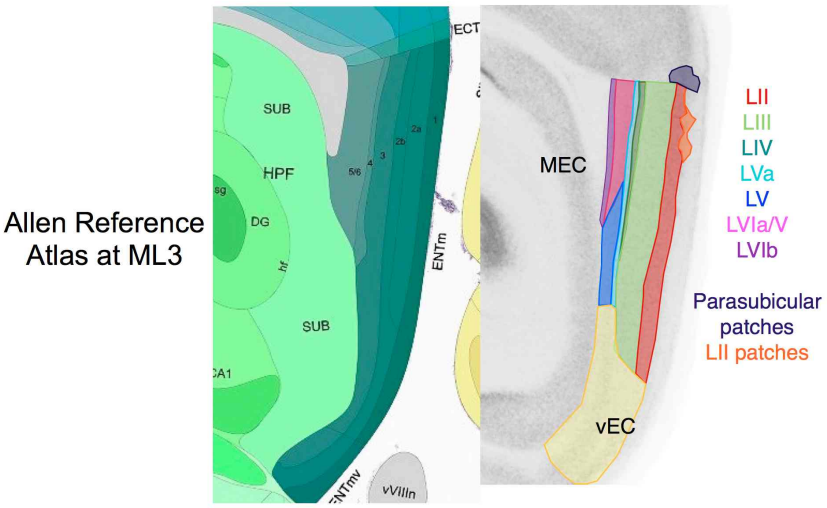


Figure 4.24: *A revised gene expression-based atlas of MEC* A refined genomic atlas of the layers and regions of MEC (right) is shown alongside the Allen Reference Atlas sagittal atlas (left). The layer V/VI region can be subdivided into a putative dorsal layer VI (and V) region and more ventral layer V region. Anterior to LVI there is a putative LVIb. Posterior to LV/VI is a putative LVa, anterior to the laminar dissecans (LIV). LII is bordered at the dorsal extent by the large patches, putative parasubiculum, and at the posterior extent by LII small patch regions, which are predominantly observed dorsally. The border between MEC and vmEC may differ across layers.

4.5 DISCUSSION

In this chapter I have investigated whether the MEC is molecularly distinct from other regions of neocortex, whether laminar differences in gene expression are organised in a functionally relevant manner and how this relates to pathology, and whether dorso-ventral differences in gene expression predict differences in gene function and dysfunction. Finally, I used the high-resolution nature of ISH data to identify putative cell types and sublayers within MEC, and to investigate the types of genes that are strongly enriched in them.

4.5.1 *Functional implications of differences between MEC and neocortical areas*

From one perspective, neurons in equivalent cortical layers but different regions perform similar functions. They receive inputs in an organised fashion based on their laminar positioning and spread of their dendritic tree, and project to other neurons within or outside the layer and to other brain structures. From another perspective, neurons in different cortical regions, even if present in the same layer, perform computations on different types of information. For these reasons, we might expect similar molecular expression in neocortical regions, to a point, but there may be important differences. Whether the MEC laminar molecular profile is similar or distinct from other cortical regions is unknown, but the organisation differs. Do organisational differences predict molecular differences?

Gene expression levels are highly correlated across the MEC, visual cortex and SS cortex, with the majority of genes showing very similar absolute expression. However, some genes show varied patterns of enrichment across the layers of the three regions, and these are far more likely to show similar laminar enrichment in SS cortex and visual cortex than in visual or SS cortex and MEC, suggesting that the molecular arrangement of MEC differs considerably from that of the neocortex. Notably, layer V/VI MEC neurons share more gene enrichment in common with layer VIa cortical neurons than layer V cortical neurons (Fig. 4.3). Further to this, genes that are enriched in MEC compared with the neocortical regions are also enriched in piriform cortex and hippocampus (Fig. 4.5). They are not enriched for particular protein attributes, but these data suggest that the computations in MEC layers that depend on molecular organisation are more different from those in visual layers than those between visual and SS layers. Differences between visual and SS cortex may arise due to the different types of inputs they are receiving and the requirement of their outputs.

4.5.2 *Layer II genes, cell type heterogeneity and dorso-ventral gradients*

MEC layer II is much more likely than the other layers to show enriched gene expression. This, in itself, was an unexpected finding; ISH data and RNA-seq data analysis have both revealed that although layer II and III of SS cortex show a high proportion of patterned genes, layer V shows the highest, and layers IV and VI show a considerable number (Belgard et al., 2011; Lein et al., 2007). This pattern not only applies to mildly enriched genes in MEC layer II, but also to genes that are unique to layer II. It suggests that many of the computations

performed in layer II are unique to this layer, whereas computations in the other layers are more generic.

Layer II genes are much more likely to be conserved than either layer III or layer V-enriched genes, or a random group of genes, but show similar conservation to uniformly enriched genes. They possess a higher proportion of homologs in Drosophila and Yeast than expected by chance. One possibility that hasn't been checked here is that layer II enriched genes and uniformly expressed genes include genes that show reasonably high expression across many cell types and brain structures, but in some cases are expressed at particularly high levels in layer II cells.

Layer II-enriched genes are overwhelming enriched for synaptic, dendritic and mitochondrial components. In line with this latter finding, many layer II-enriched genes are involved in mitochondrial pathways such as oxidative phosphorylation, but only if they also show a gradient in expression pattern. The enrichment of mitochondrial genes may signify that dorsal layer II cells have the largest energy requirement in the MEC. It may also be related to the high proportion of conserved genes. Relevant to energy requirement, dorsal cells are also more likely than ventral layer II cells to be enriched for immediate early genes, which are indicators of cellular activities.

Enrichment of immediate early genes, plasticity-related genes, synaptic and dendritic genes, and enrichment of the LTP pathway (predominantly post-synaptic), most of which only apply to dorso-ventrally organised of genes with no gradient, may all be indicators that learning mechanisms in MEC heavily depend on plasticity onto layer II cell synapses. At the other end of the neuron, the enrichment of axon guidance pathway molecules may reflect the demand on layer II neurons to form new functional connections onto newly born granule cells dendrites. Axon guidance molecules are also dorso-ventrally organised, but it is not known whether dorso-ventral differences in connectivity between layer II and DG neurons exist.

Patterns of expression that mark out subsets of cells in MEC layer II are common. First, the large patches dorsal to layer II have not been universally defined (Boccara et al., 2010), but Burgalossi et al. (2011) have posited that they are a continuum of parasubiculum in that the large patch cells are smaller than MEC cells and they interconnect with other parasubicular regions. The idea that the large patches are a distinct region from MEC is supported by the fact there is a clear gene expression border at the dorsal extent of MEC layer II where it meets the patches (see Fig. 4.19). Moreover, large-patch genes are also expressed in the region of parasubiculum medial to MEC that is considered superficial parasubiculum (Caballero-Bleda and Witter, 1993). There are also many examples where large-patch genes are expressed in cells along the layer I/II border of the remainder of MEC. It remains to be clarified whether these make up the same cell type.

The identification of these cells may be important because large-patch neurons, like grid cells, are theta-modulated (Burgalossi et al., 2011), and as they will be encountered as tetrodes are lowered, they are likely to be included in grid recordings without necessarily being identified as non-MEC. Their behaviour is different from MEC neurons, however; they show strong head-direction selectivity and fire during the descending phase (Burgalossi et al., 2011). Recent work that carefully examined dorso-ventral changes in grid cell properties (Stensola et al., 2012) did not discuss the possible difference.

Genes that are specifically expressed in layer II cells are also enriched for the development of neuron projections function. Do these genes define the cells that project to the dentate gyrus? Layer II contains at least two cell populations that have already been defined (Varga et al., 2010); the calcium-binding protein calbindin is expressed in cells that form small patches invading layer I while reelin appears to avoid these small patches. While there is evidence that calbindin neurons project to regions other than the hippocampus, reelin neurons project to the dentate gyrus (Varga et al., 2010). I could subdivide genes expressed in layer II into three groups, independent of their dorso-ventral patterning. Small patches genes show a similar, but not identical, expression to calbindin, a further subgroup avoid the patches in the same way as reelin, and a third group are expressed densely or sparsely across both regions. Cholecystokinin is the only axonogenesis-related molecule in the patches group, but a significant proportion of 'patches' genes are involved in synaptic transmission, which may be functionally relevant. The current analysis was not sensitive to identifying Reln-like patterns in gene expression, but it would be interesting to determine whether particular axon guidance or cell adhesion molecules follow a similar pattern of expression. Double *in situ hybridisation* experiments investigating whether patchy genes and Reln-like genes indeed mark out different populations of cells in the same section containing layer II could confirm that gene expression distinguishes cell type.

The expression of post-synaptic density genes also appears to distinguish dorso-ventral layer II genes from genes that are absent in the very most dorsal region, which can also be detected using ISH, and genes that show ventro-dorsal layer II expression, suggesting that the PSD variability could be an important determinant of differences along the layer II gradient.

4.5.3 *Layer III genes*

Genes are very rarely enriched in layer III of MEC, compared with layers II and V/VI. However, a large proportion of genes expressed in layer II are also expressed in layer III, but not in layer V/VI, suggesting that layers II and III have similar dependencies on certain genes. Layer III contains a multitude of cell types, including a large population of pyramidal neurons (summarised in Canto et al., 2008) and directly receives many of the inputs from other regions, including the parasubiculum and presubiculum (Canto et al., 2012), which target the dendrites of all three layers. The input from layer III to CA1 is the only pathway that has been shown to be necessary for place field stability (Brun et al., 2008a).

Genes that are unique to layer III are actually underrepresented for homologs in yeast (Fig.4.10), which suggests that they are recently evolved genes. The low number means that it is difficult to detect functional enrichment, but the enriched genes include a number of ion channels and calcium-binding proteins. Aside from the role in the hippocampal input, layer III neurons have also been implicated in epilepsy (Kumar and Buckmaster, 2006). However, there is no clear overrepresentation of epilepsy genes amongst layer III enriched genes. Layer II neurons also show abnormalities in epilepsy, so it is possible that genes shared between layers II and III contribute towards vulnerability.

4.5.4 *Layer V/VI genes*

Layer V and VI are the deep layers of MEC that predominantly receive hippocampal input rather than projecting to the hippocampus (Swanson and Cowan, 1977; Tamamaki and Nojyo, 1995). As such, it might be predicted that genetic differences between layer V/VI and layers II and III are greater than within these layers. Layers V/VI are enriched for fewer genes than layer II, but more than layer III. They share a considerable number of genes with layer II, but very rarely are genes enriched in layers III and V, but not layer II. Layer V/VI-enriched genes also appear to exhibit the most similar expression patterns in neocortex out of all the enriched genes.

Like layer III, genes enriched in layer V/VI are actually less likely than chance to be conserved in yeast, suggesting that it contains more recently evolved genes. The genes found here to be enriched in layer V/VI are not uniquely enriched for any particular protein attributes, but do show high overrepresentation of genes involved in transcriptional regulation. There is some evidence that genes specific to oligodendrocytes (Cahoy et al., 2008) are more enriched in layer V/VI, particularly at the dorsal extent, than other layers. It was recently shown, also using the ABA dataset, that oligodendrocyte-specific gene expression distinguishes regions, but not to same extent as neuron-specific genes (Ko et al., 2013).

In this research I've combined layers V and VI on the basis that in the sagittal plane they cannot be clearly distinguished anatomically. It is possible to use ISH data to identify genes that characterise very narrow regions of MEC, including a region corresponding to layer Va and to a layer between layer V/VI and the hippocampus, which I have putatively classed as layer VIb, as it is continuous with layer VIb of the neocortex. Many of the genes in layer Va or VIb are expressed in both these layers but not dorsal layer V/VI proper, and they are overrepresented for post-synaptic density genes. In this thesis I haven't directly addressed if gene expression distinguishes layer V from layer VI, but the apparent dominant ventro-dorsal gradient in layers V/VI appears to provide a clue. Data from the coronal plane and horizontal plane, if available, may clarify the possibility that layer V occupies the more ventral MEC region, as viewed sagittally, whereas layer VI invades the deep region of MEC. Since the border between layers V and VI at a given point along the dorso-ventral axis is most clearly visible in horizontal sections, it will be important to clarify that genes that appear to be ventro-dorsally organised, i.e. specific to layer V, show this pattern of expression in the horizontal plane.

4.5.5 *Gene expression in MEC and disease*

4.5.5.1 *Layer II dorsal-ventrally organised genes and Alzheimer's Disease*

The MEC, specifically layer II, is one of the first regions of the brain to show pathology in brains of patients suffering from Alzheimer's disease (Braak and Braak, 1991), a disease that has been linked to abnormalities in a number of different genes (see (Bertram and Tanzi, 2008) for a review). There is evidence that the cytochrome oxidase (CO)-rich patches of MEC layer II are destroyed in AD (Solodkin et al., 1996). If patchy organisation is particularly important

to spatial processing, this could explain why loss of spatial orientation is an early clinical symptom in patients (Cherrier et al., 2001; Burgalossi et al., 2011). I was therefore particularly interested in examining whether genes linked to AD, or that encode proteins disrupted in the disease, are particularly enriched in MEC or its layers, specifically layer II.

It emerges that genes involved in the AD pathway, as defined by KEGG (Kanehisa et al., 2004), are actually expressed at similar levels across MEC as in visual and somatosensory cortices (Fig.4.7), and that genes enriched in MEC, relative to neocortical regions, show no specific enrichment for AD. However, approximately one quarter of the genes involved in the AD, PD and HD pathways are enriched in MEC layer II compared with other layers. Although genes expressed at higher levels in dorsal regions are also extremely enriched in the neurodegenerative disease pathways, whereas ventro-dorsal genes and genes showing no gradient are not, this enrichment is likely due to the fact that many dorso-ventrally organised genes are also enriched in layer II. These results suggest that clues as to the vulnerability of MEC to neurodegeneration are most likely to be found in the activity of layer II cells, particularly at the dorsal extent. These cells are also enriched for genes involved in mitochondrial function, which is disrupted in Parkinson's disease (Abou-Sleiman, as cited in (Belgard et al., 2011)) and Alzheimer's disease (reviewed in Schon and Area-Gomez, 2013). This could explain the disruption to CO activity (Solodkin et al., 1996) in MEC.

As a cautionary note, it should be highlighted that many of the proteins involved in the AD and neurodegenerative pathways may not be specific to these diseases but show either disruption in humans or animal models, or are targeted by drugs in these diseases. It is not necessarily the case that genes that, when mutated, increase risks of specific diseases are enriched in MEC layer II. Nevertheless, the vulnerability of MEC layer II to pathology could result from the enrichment of a significant number of elements of the AD pathway that are disrupted by genetic or environmental factors.

4.5.5.2 *Epilepsy, schizophrenia and autism*

Cells across both layers II and III demonstrate abnormalities associated with epilepsy (Kumar and Buckmaster, 2006; Du et al., 1993), but I found that only layer II-enriched genes show enrichment for epilepsy-related genes. Layer II genes also include a high proportion of autism-related genes, whereas schizophrenia-related genes are overrepresented across all the layer-enriched genes, but not uniformly expressed genes. Schizophrenia genes are most over-represented amongst layer-V/VI specific genes, but to my knowledge layer V/VI neurons have not been shown to demonstrate specific abnormalities. Ventro-dorsal genes and genes with no gradient are more likely than dorso-ventral genes to be involved in schizophrenia and autism, suggesting that it is the ventral layer V/VI region that shows enrichment. These genes are particularly enriched for synaptic transmission, despite the small size of the list, which is thought to be disrupted in schizophrenia (summarised in Arnold, 2000). It is possible that enrichment only exists because of the link between these genes and synaptic transmission. Given that the genetic links and disrupted pathways of these diseases are not well understood, it is difficult to draw strong conclusions from these results, but a combination of disease-related genes enriched in specific areas of MEC could contribute to vulnerability of MEC cells.

4.5.6 *ISH as a tool for improving understanding of neural circuits*

In this research I used *in situ hybridisation* for several different types of analysis: to quantify gene expression within regions of interest for the purpose of analysing differences in protein function, and to use patterns in gene expression to understand the distribution of neurons expressing different genes. Although the ABA data has been extremely useful for the former task, its reliability must be questioned given the small number of mice in which each gene has been measures (often just 1). This is particularly true for the genes with more generic patterns of RNA labelling, where the specificity of the probe is not immediately evident. In addition, in this chapter I have made comparisons between gene expression in different brain regions on the assumption that probe binding efficiency is equivalent, but it is possible that this is not the case.

By contrast, the highly specific and anatomically-related patterns that ISH reveals provides strong support for the use of ISH to identify genes with highly specific expression. Adding to this is the fact that the Allen Institute have validated the 1000 genes that are most frequently accessed using their web tool, using literature search and results from other online brain atlases of gene expression. This suggests that a significant amount of their data is reproducible. Highly specific patterns in expression such as those shown here will enable us to build up a more complete view of the anatomy of MEC that incorporates both local and distant connectivity, and molecular profile. A three-dimensional atlas at a higher resolution than provided by the ABA could significantly facilitate the task of researchers attempting to understand the organisation of MEC, as well as other brain regions.

4.5.7 *Future Work*

This type of analysis of ISH data can be taken in many further directions. The most useful possible outcome would be to show that genes that appear to characterise specific cell types or layers in fact provide molecular signatures that map on to MEC function. This will require both validation of their pattern of expression using ISH or immunohistochemistry and functional analysis of the contribution of these genes. It remains unclear in MEC which cell profile is even the most interesting to study. Work by (Garden et al., 2008) shows that synaptic integrative properties of cells that display a prominent sag in their electrophysiological profile are tuned along the dorso-ventral gradient. By combining electrophysiological recordings and ISH or immunolabelling of the patchy genes that show gradients similar to the Hcn1 channel (which controls the sag), it might be possible to elucidate the molecular profile of these cells and identify the best candidate genes for manipulating them.

An alternative direction for further study is to query the system-level function of immediate early genes and plasticity-related genes that show a dorso-ventral gradient and are enriched in layer II of MEC. This would not only be informative about the type of coding and learning that cells along the gradient undergo, but could also provide insight into why genes involved in disease pathways show similar expression patterns.

The analysis could also benefit from several improvements that would provide improved anatomical information. For one, the rostro-caudal and medio-lateral positioning of MEC and

its topographical mapping is less well established than that of the dorso-ventral positioning, particularly in terms of individual layers. I showed in the previous chapter that relative gene expression in the three major cell body layers is highly correlated in ML2 and ML3, but not all genes show the same relative laminar expression along this axis. It would be interesting to consider the patterns of change across the medio-lateral extent in more detail and to determine whether differences reflect continuous changes in gene expression within a particular cell type or a change in the distribution of the cell type. A small number of genes show noticeable changes in expression from sections corresponding to ML1 to sections corresponding to ML4. For example, the gene Endophilin-A1 (*Sh3gl2*) shows very high expression in layer V/VI in ML1, but more medially, layer V/VI expression reduces, particularly in more dorsal regions, relative layer II expression increases and ventral layer III expression decreases (Fig. 4.25(A)). As yet, there are no anatomical or physiological explanations for these changes.

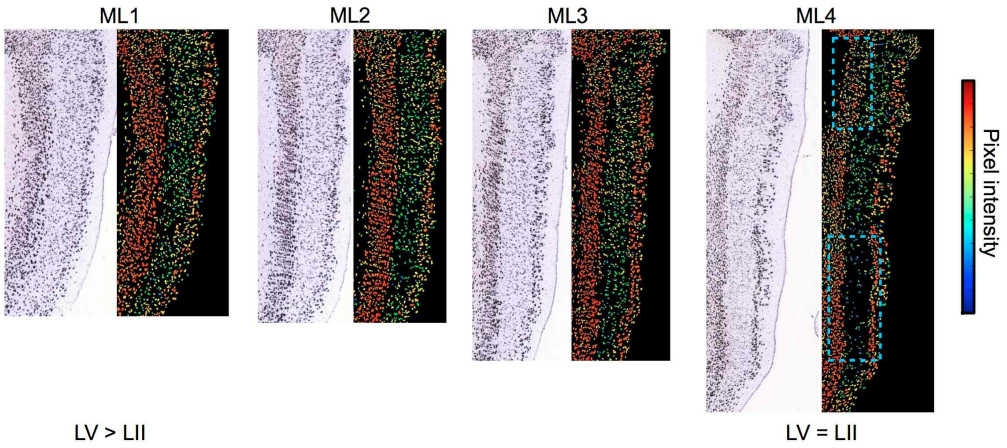


Figure 4.25: *Medio-lateral changes in gene expression.* ISH and expression level images showing the changes in gene expression from ML1 to ML4 for an example gene Sh3gl2. Blue boxes indicate regions in ML4 showing the highest difference to ML1.

4.5.8 *Conclusions*

Gene expression is highly variable across MEC, often with patterns that directly correspond to known functional distinctions and pathology. I have identified a number of genes, possibly the only genes, that show highly selective expression in particular layers of MEC. I have also identified genes that very selectively mark dorsal or ventral regions of layers, and sub-layers, which may be useful for researchers wanting to study the significance of such gradients to behaviour. One particularly interesting observation is the dorso-ventral organisation of immediate early genes, which are indicators of activity, and can also be indicators of potential learning (Flavell and Greenberg, 2008). Immediate early genes such as the activity-regulated cytoskeleton-associated gene, Arc, could be useful indicators of the distribution of activity and plasticity across the MEC circuit.

5

A DETAILED ANALYSIS OF BASELINE ARC EXPRESSION IN MEC

5.1 SUMMARY

In the previous chapter I identified a number of genes of functional interest that show specific patterns of expression in MEC. The gene encoding the activity-regulated cytoskeleton-associated protein (Arc) is expressed at high levels in MEC, and shows a dorso-ventral gradient in RNA expression in *in situ hybridisation* data from the Allen Brain Atlas, which suggests that it may be of interest for understanding the physiological function of MEC. Here, I provide a detailed examination of basal Arc expression in MEC, in preparation for subsequent analysis of activity-dependent changes in the following chapters. To estimate the distribution of MEC cells that express Arc at baseline, I have quantified direct- and immuno-fluorescence of GFP in heterozygous Arc^{+}/GFP^{+} mice that express GFP under the control of the endogenous Arc promoter and compared this with protein data. The results are summarised below:

1. Baseline Arc-GFP Arc activation is low in MEC. This suggests that typical environmental cage conditions do not strongly drive Arc-GFP expression in MEC.
2. Arc-GFP activation follows a gradient in MEC layer II under resting conditions. Arc may therefore be of interest for understanding the development and maintenance of other physiological gradients.
3. Arc-GFP activation is higher in superficial cortex than MEC at baseline, but CA1 expression is similarly low.
4. Baseline expression of Arc protein is also low in MEC but shows subtle differences in its somatic expression between the superficial and deep layers. This suggests that these circuits could be differentially engaged at rest.

5.2 INTRODUCTION

Activity of neurons in the medial entorhinal cortex (MEC) encodes an animal's position in its environment. A gradient in the resolution of this spatial code along the dorsal-ventral axis (Hafting et al., 2005; Stensola et al., 2012) is mirrored by tuning of the integrative properties of MEC neurons (Garden et al., 2008). A central aim is to show how MEC function, and in particular this gradient, is maintained and controlled. As a starting point, I demonstrated in the previous chapter that over 2000 genes are enriched in different MEC layers, and over 1000 exhibit a gradient in expression.

The gene that encodes the activity-regulated cytoskeleton-associated protein (Arc) is one such gene, exhibiting higher levels in superficial than deep layers of MEC and following a dorso-ventral gradient (see Fig. 5.1(a)). Arc is an immediate early gene that is essential for certain forms of plasticity (Plath et al., 2006), homeostasis (Shepherd et al., 2006) and spatial learning (Plath et al., 2006), and therefore could be of considerable interest to understanding the coding of information in MEC. Previous work has shown that Arc is expressed at low basal levels in hippocampus (Messaoudi et al., 2007) (Fig. 5.1(b)) and visual cortex (Wang et al., 2006) (Fig. 5.1(c)), and at high levels in areas such as the reticular hypothalamic nucleus (Ons et al., 2004).

Despite considerable investigation of Arc function in hippocampal tissue (see Messaoudi et al., 2007; Ons et al., 2004), there is very little work examining MEC. As such, the relationship between the activity of MEC neurons, their integrative properties, activity-dependent gene expression and plasticity mechanisms is not understood. In this chapter my goal is to provide a detailed characterisation of the baseline expression of Arc in animals taken directly from their undisturbed home cage and sacrificed. This provides a starting point for characterising how Arc expression changes with activity.

In order to provide a comprehensive analysis of the baseline activity of Arc across cells, I have made use of heterozygous Arc-GFP mice that express an enhanced GFP construct, d2EGFP, under the control of the endogenous Arc promoter (Wang et al., 2006). These mice provide two complementary measures of the distribution of Arc activity across cells: Arc protein expression and the more easily visualisable somatic GFP expression, which show highly overlapping novel exploration-induced expression patterns in visual cortex (Wang et al., 2006). A possible limitation of measuring Arc based on quantifying protein level using immunofluorescence is that Arc can be translated when transcription is blocked (Waung et al., 2008) and there is evidence that Arc mRNA is trafficked from the soma without having been translated (Bramham et al., 2010; Giorgi et al., 2007). Given that dendrites of cells can branch and extend over long distances, somatic Arc protein does not necessarily therefore provide a direct indicator of the distribution of Arc promoter activation across cells. In Arc-GFP heterozygote mice Arc promoter activity drives the transcription of GFP mRNA and Arc mRNA, which share the 5′ UTR and ATG start codon regions (Wang et al., 2006). Given that the regions of the Arc mRNA that enable trafficking are in the coding and 3′ UTR regions (Kobayashi et al., 2005), GFP mRNA is unlikely to be trafficked and will instead be translated and accumulate in the soma. As such, somatic GFP likely provides a close indicator of the distribution of Arc promoter activity across cells. The observation that certain regulatory proteins that bind

to the shared 5′ UTR of GFP and Arc only suppress Arc translation in the dendrites (Niere et al., 2012) provides further support for this, although the requirement to translate GFP could limit its ability to directly report promoter activity. Importantly, these mice show no overt behavioural abnormality, but previous work has shown that they have deficits in visual cortex plasticity and orientation specificity (Wang et al., 2006). When maintained in darkness they express very low levels of GFP in visual cortex (see Fig. 5.1(c)) (Wang et al., 2006).

To provide a high-quality quantitative analysis of expression, I have used high-resolution confocal stack images combined with a thorough cell-based analysis of expression. I adapted the technique that was previously developed for analysing the native GFP fluorescence signal from this mouse line (Wang et al., 2006). This allowed me to analyse both direct GFP fluorescence and GFP antibody labelling across all defined neurons. The reason for using both measures is to preserve the most direct measure of expression, free from potentially confounding immunohistochemical factors, while simultaneously being able to monitor the density and cellular localisation of expression, and achieve a higher sensitivity.

My prediction is that, similar to other brain regions, Arc expression will be relatively low in MEC. The MEC is considered to be a region important for spatial coding (Hafting et al., 2005), so these circuits are unlikely to be highly active when mice are within their home-cage. However, there may be subtle differences between MEC, hippocampus and visual cortex if they are differentially recruited under the 'baseline' conditions employed here.

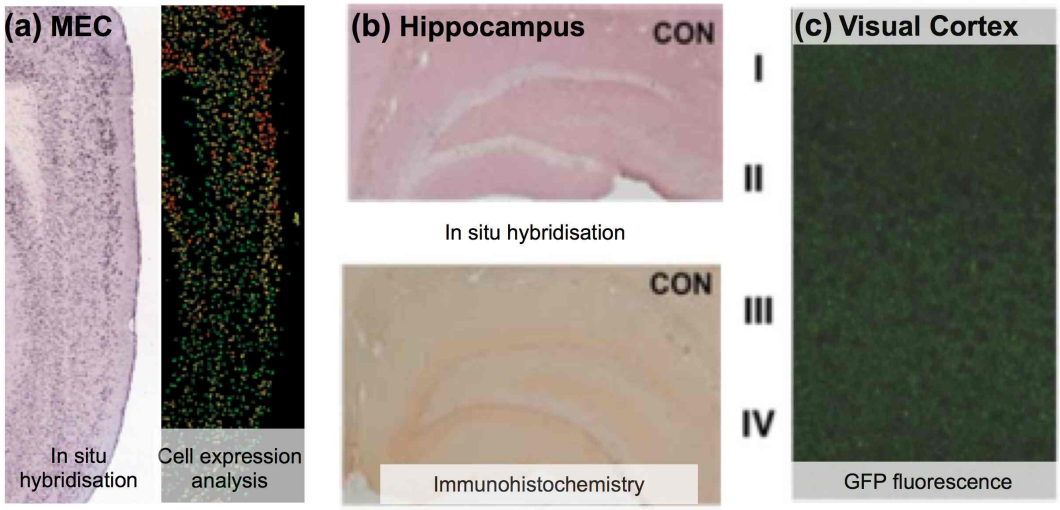


Figure 5.1: *Baseline expression of the activity-regulated cytoskeleton-associated gene (Arc).* (a) Images from the Allen Brain Atlas resource showing *in situ hybridisation* (left) and a cell-based expression level analysis (right) for Arc in a 56-day-old C57Bl/6J mouse. Arc mRNA is expressed with a dorso-ventral gradient in MEC. (b) Brightfield images showing Arc *in situ hybridisation* (upper) and immunohistochemistry (lower) in the rat dorso-medial hippocampus under baseline conditions (from (Messaoudi et al., 2007)). (c) Confocal image showing directly imaged GFP protein expression (where GFP is under the control of the endogenous Arc promoter (Wang et al., 2006)) in the visual cortex of heterozygous Arc^{+}/GFP^{+} mice sacrificed under dark conditions. Arc is expressed at low levels under such conditions (Wang et al., 2006).

5.2.1 *Aims*

- To examine basal Arc expression (using GFP as a marker) in MEC. This will demonstrate whether Arc is activated in MEC cells in the absence of a specific task. Since Arc is a marker of neuronal activity and is involved in plasticity, this will provide insight into the baseline activity of the MEC circuit and whether there are laminar differences.
- To investigate the presence of a gradient in GFP or Arc protein expression in layer II. This may provide insight into differences in synaptic properties and plasticity mechanisms between dorsal and ventral cells that map onto other physiological and behaviourally relevant gradients.
- To compare MEC expression of Arc with that of visual cortex, which has been investigated previously (Wang et al., 2006), and with area CA1, areas that show extensive connectivity with MEC. This will not only provide a means of validating the results presented here for MEC, but will enable a comparison of baseline activity levels in these regions with that of the MEC circuits.
- To compare GFP and Arc protein expression to provide an indication of the validity of using these Arc-GFP mice as reporters.
- To test whether similar patterns in basal Arc -GFP activation are present in Arc^{+}/GFP^{+} and GFP^{+}/GFP^{+} (Arc knockout) mice. There is evidence that the downstream effects of *Arc* activity regulate its subsequent levels (Bramham et al., 2010), so it will be interesting to test whether there are differences in the pattern of expression.

5.3 METHODS

5.3.1 *Development and validation of the Arc-GFP mice used to examine Arc expression*

C57BL/6-Arctm1Stl/J mice (Jackson Labs 007662) (Wang et al., 2006), hereafter 'Arc-GFP mice', were used to examine Arc expression. In these mice an altered GFP construct (d2EGFP) is expressed under the control of the endogenous Arc promoter (Fig. 5.2(A)(i)). This particular alteration shortens the half-life of EGFP expression to 2 hours (Fig. 5.2(A)(ii)), comparable to the decay of Arc, as measured by the fluorescence intensity of cells treated with the protein synthesis inhibitor cycloheximide (Living Colors^{TM} Destabilized EGFP Vectors (1998) CLONTECHniques XIII(2):16–17.).

5.3.1.1 *Method for the development of Arc-GFP mice (Wang et al., 2006)*

The Arc-GFP mice were developed by Wang et al. (2006) by first cloning a 5.5 kb EcoRI/AflII segment of Arc genomic DNA from a C57/BL6 mouse BAC library (Genomic Systems). In the targeting vector, d2EGFP (Clontech) is followed by a selectable marker in the form of a floxed neomycin resistance cassette (in the same orientation) into the Nde1 site after the ATG start codon of the gene (See Fig 5.2(A)(i) for a schematic). This vector plasmid was electroporated into C57BL/6-derived embryonic stem (ES) cells. The first 48 base pairs of the open reading frame before the SacII site were deleted as a result of recombination. ES cell clones were injected into BALB/c blastocysts to derive chimeric mice. Germline transmissions of targeting constructs in C57/BL6 mice were identified by coat color and confirmed with Southern blot and PCR. This strain has been maintained on the C57BL/6J background.

5.3.1.2 *Validation of Arc expression in heterozygous and homozygous Arc-GFP mice by Wang et al. (2006)*

Using Western Blot and an Infrared western blot imaging system (LICOR), Wang et al. (2006) showed that heterozygous Arc-GFP (Arc^+/GFP^+) mice express approximately 80% of the level of Arc protein across the whole brain as their wild-type (WT) littermates, whilst their GFP^+/GFP^+ littermates express no Arc protein (Fig. 5.2(A)(iii)). Arc protein level was normalised against that of the housekeeping enzyme GAPDH. Wang et al. (2006) also developed GFP and Arc primers for RT-PCR to test RNA expression. Arc^+/GFP^+ animals express full-length Arc RNA, as shown using primers for the 5' and 3' regions (Fig. 5.2(A)(iv)). The Arc truncated open reading frame (ORF) appears to be expressed at similar levels between Arc^+/GFP^+ and WT animals whereas WTs show higher levels of the initial region of Arc RNA, corresponding to the site of the GFP insertion in Arc^+/GFP^+ mice. Conservation of this Arc 3' UTR could, in theory, enable the translocation of GFP RNA to dendrites by similar mechanisms to that of Arc (Kobayashi et al., 2005), but its absence from GFP^+/GFP^+ animals suggests that it is not maintained in GFP mRNA (See Fig. 5.2(A)(iv)).

5.3.2 *Maintenance and identification of the Arc-GFP mice*

5.3.2.1 *Husbandry and Breeding*

We purchased Arc-GFP mice from Jackson Labs and they were provided as heterozygous Arc^{+}/GFP^{+}. All mice used in this experiment were bred in-house from Arc^{+}/GFP^{+} pairs. No external animals were introduced into the breeding scheme. All mice used were male and 7-11 weeks old. Mice were fed *ad libitum* and were housed with a minimum of two per cage on a standard light cycle of 14 hours light, 10 hours dark (7am-9pm).

5.3.2.2 *Confirmation of genotype*

Arc^{+}/GFP^{+} mice, which express both a copy of normal Arc and a copy of the GFP transgene, were used in the majority of experiments (except where explicitly specified). All animals used were ear-punched and genotyped prior to the experiment. In the case of ambiguity over genotype, post-mortem collected tail clips were used to confirm genotype. Two separate PCR reactions were run; one containing mutant primers, and the other WT primers (See Fig.5.2(B)(i-ii) for primer targets and sequences). Primer sequences were provided by Jackson Labs and were purchased from Source Biosciences. The PCR reaction was adapted from that provided by Jackson labs. Briefly, DNA was extracted from ear punches using 20 µl DNAReleasy. PCR reactions were made up to 12µl for each DNA sample, which included 6µl GoTaqGreen 2x MasterMix, 0.2µl DNA, 1.66µl betaine, 0.075µl each of Beta-globin forward and reverse primers, 0.1µl of either WT or mutant forward and reverse primers, and 3.79µl Sigma nuclease-free water (see Appendix A for full protocols). All Arc^{+}/GFP^{+} mice were confirmed to be positive for both WT and mutant DNA based on the presence of a dark band at 186bp (WT) and 250bp (Mutant) (Fig.5.2(B)(iii)). GFP^{+}/GFP^{+} mice expressing two copies of the GFP transgene (Arc knockout animals) were also used in this experiment to investigate the pattern of expression in the absence of Arc protein. The genotype of GFP^{+}/GFP^{+} mice was confirmed based on the absence of a band at 186 bp in the WT primer reaction.

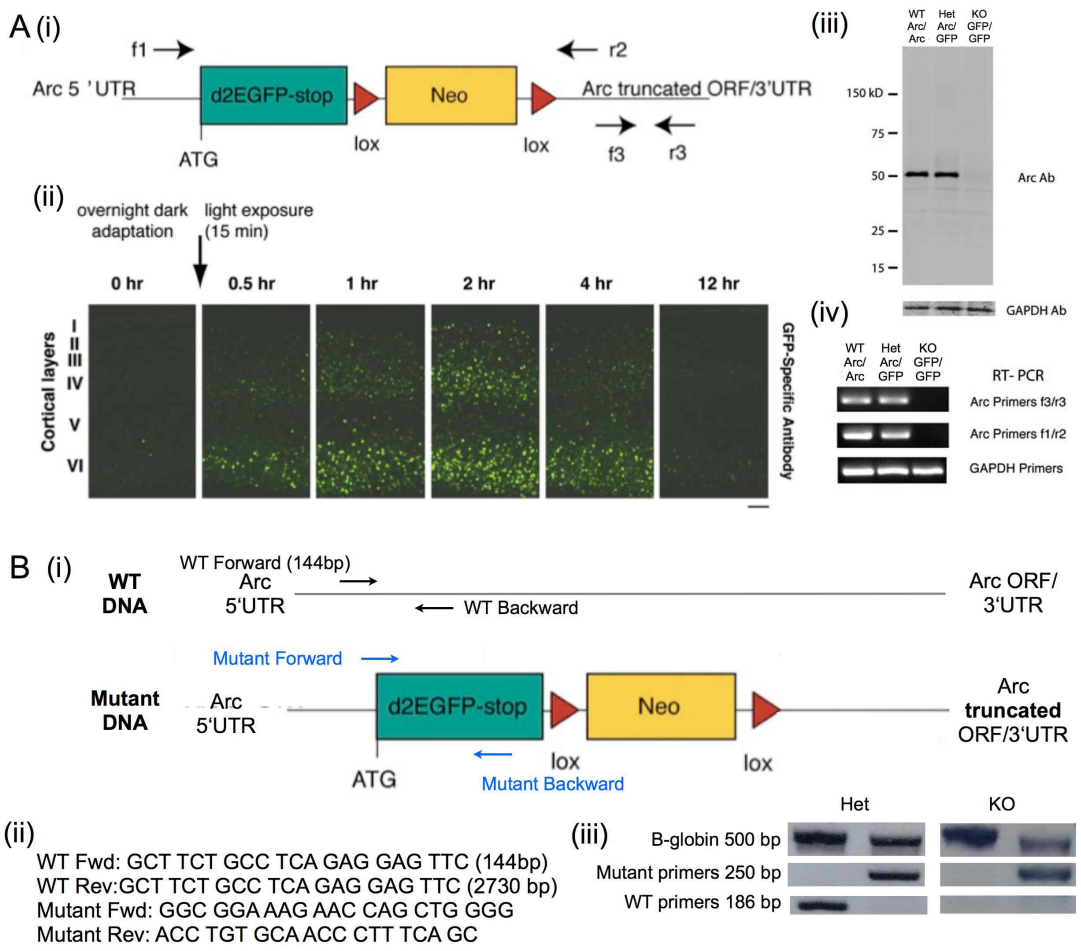


Figure 5.2: *Characterisation of RNA and protein expression in Arc-GFP mice* (A) From Wang et al. (2006) Fig.1: (i) Schematic showing the positioning of the enhanced GFP construct (d2EGFP) under the control of the endogenous Arc promoter. The GFP construct is followed by a Neo cassette and the Arc truncated ORF and 3' UTR. Arrows and labels mark the positions of the primers used in the RT-PCR reaction. (ii) Example confocal images of visual cortex showing the time course of GFP expression (detected using a GFP-specific antibody) following exposure of a mouse to a novel environment (Wang et al., 2006). GFP protein peaks 2 hours following activity (exposure to a novel environment under light conditions) and returns to baseline within 12 hours. (ii) Example results from a Western Blot with an antibody against Arc (Santa cruz: H300) comparing whole-brain tissue from WT (left), Arc^+/GFP^+ (middle) and GFP^+/GFP^+ mice. Arc^+/GFP^+ and WT mice show a positive band at 50kDa, the size of protein, whereas GFP^+/GFP^+ mice express no Arc protein. All mice were positive for the control GAPDH antibody. (iv) Example of results from RT-PCR reactions for WT Arc^+/Arc^+(left), het Arc^+/GFP^+ (centre) and KO GFP^+/GFP^+ (right) animals. For Arc primers in the truncated region (f3/r3)(upper) and across the insert (f1/r2)(mid) regions, positive results were only found in WT and Arc^+/GFP^+ mice, while primers for a control housekeeping gene GAPDH yielded positive results across all mice. (B) Current data: (i) Schematic showing the position of the WT (black arrow) and mutant primers (blue arrow) on the WT (upper) and mutant (lower) DNA. (ii) Primer sequences for WT and mutant forward and reverse reactions. (iii) Example DNA product from a PCR reaction containing WT (upper) and mutant (lower) primers for Arc^+/GFP^+ (left) and GFP^+/GFP^+ mice (right) used in this experiment. All Arc^+/GFP^+ animals exhibited an upper band, revealing the presence of the mutant GFP DNA product (250bp), and a lower band, demonstrating the presence of WT DNA product (186bp), whereas GFP^+/GFP^+ mice showed only an upper band. All reactions were performed in the presence of control Beta-globin primers such that erroneous results could be detected by the absence of a band at 500bp.

5.3.2.3 *Differences in RNA and protein across Arc-GFP mice*

In WT Arc^+/Arc^+ and Arc^+ /GFP^+ mice, the Arc gene is transcribed to mRNA and translated into Arc protein (See Fig. 5.3). Post-translational and other regulatory mechanisms may regulate both levels of Arc mRNA and protein, and may have different effects in WT and Arc^+ /GFP^+ animals due to differences in the levels of Arc protein. Arc^+ /GFP^+ and GFP^+ /GFP^+ animals both express the GFP transgene and therefore express GFP protein. However, the levels of GFP protein are likely to differ due to differences in copy number.

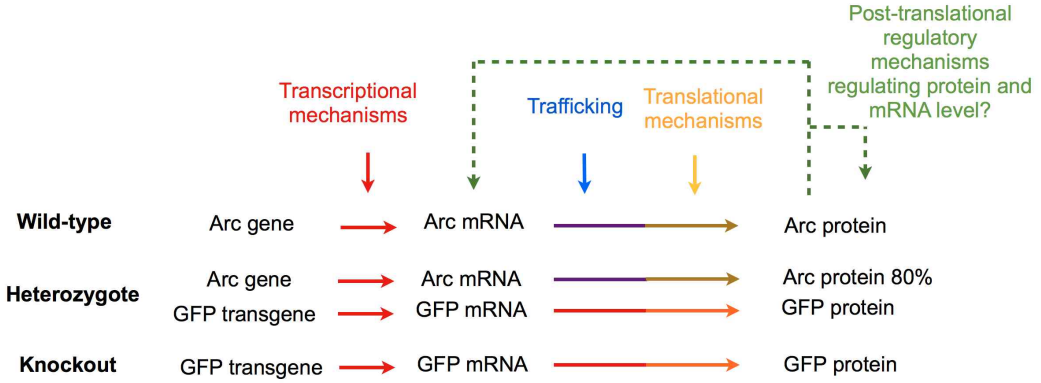


Figure 5.3: *Summary of differences in Arc and GFP protein expression in Arc-GFP mice with different genotypes* Schematic showing expected differences at the level of RNA and somatic protein in WT, Arc^+/GFP^+ and GFP^+/GFP^+ animals. Differences that are difficult to predict may be caused by trafficking, post-translational regulatory mechanisms that differentially affect mRNA and protein levels depending on their current expression.

5.3.3 *Behavioural Procedure and Sacrifice*

To ensure that Arc expression could be measured at its baseline in MEC, mice were housed in a room separate from the main holding rooms of the animal facility for at least seven days prior to the experiment. This timing ensured that they were habituated to the room by the time of the experiment. There was no entry to this room other than by technicians for daily checks and the room was not disturbed within 20 hours prior to the experiment. Mice were not explicitly handled prior to the experiment. When handling was necessary, mice were always picked up by the tail before being immediately returned to a solid surface.

On each experiment day between 8.30am and 9.30am (light cycle onset at 7am) one animal was removed from its cage and immediately anaesthetised with 100% isofluorane and sodium pentobarbital (0.1-0.12ml, or Euthatal 0.05ml) before being transcardially perfused with approximately 25ml 0.9% saline and 20ml 4% methanol-free paraformaldehyde (PFA) solution. See Fig 5.4(i-iii) for a summary of this process. The saline step ensured maximum possible blood removal to provide minimum background staining for immunofluorescence. The PFA stage was monitored for the presence of a characteristic tail dance, and appropriate fixation was confirmed by stiffness of the tail. Experiments were carried out across 4 separate sessions. Efforts were made to ensure these were as similar as possible.

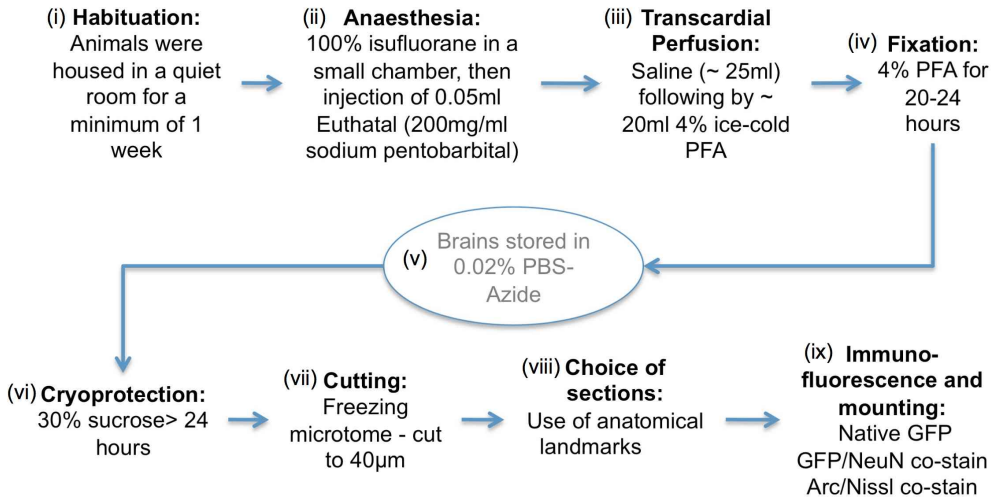


Figure 5.4: *The procedures for animal treatment and tissue processing.* Outline shows each stage of processing from initial habituation (i) to mounting of sections (ix). This was performed identically for each animal except for some variation in (v) the holding time in which sections were stored in PBS-Azide (white oval), which lasted between hours and months, but was always equivalent across pairs of mice.

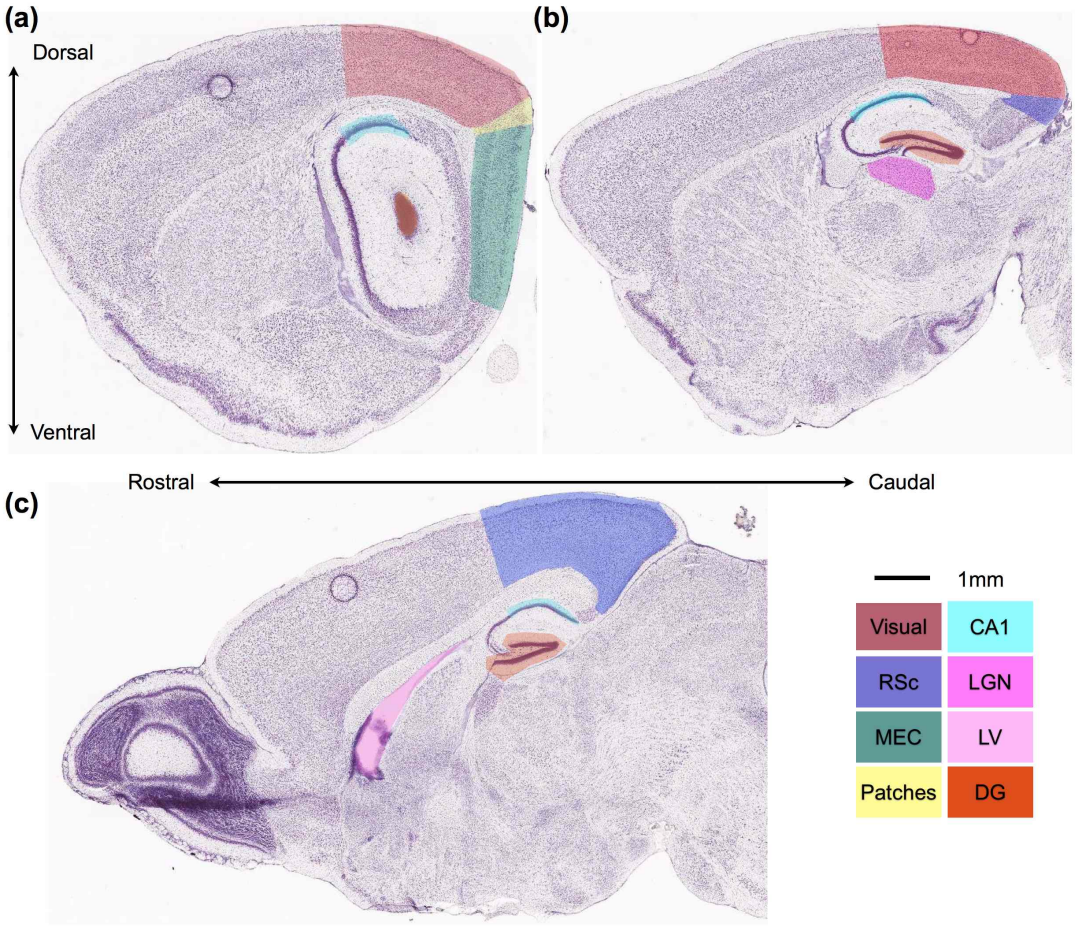


Figure 5.5: *Important landmarks for identifying sections of interest* Example images from the Allen Brain Atlas developmental reference atlas showing Nissl-stained sagittal sections from 56-day old mice that have been colour-coded to show the anatomical markers used to identify sections to be processed for immunohistochemistry and imaging. (a) Sections including the MEC (green) and visual cortex (red) were chosen at the medio-lateral extent at the point where the granule layer of dentate gyrus (orange) first appears as a small oval in the centre of the hippocampal region. This is approximately 3.1mm from the midline (assuming a Bregma-Lambda distance of 4.2mm typical of an adult male C57Bl/6 mouse at 56 weeks). (b) Sections containing CA1 (cyan) were chosen at approximately 2.2mm lateral to the midline, which can be identified by the shape of the hippocampus and presence of the dorsolateral geniculate nucleus (magenta) just ventral to the dorsal hippocampus. (c) Sections including the dentate gyrus (orange) were chosen at the point where the lateral ventricle (magenta) is elongated in the rostro-caudal plane.

5.3.4 *Tissue Preparation and Immunohistochemistry*

Brains were stored in 4% PFA for 20-24 hours (Fig 5.4(iv)). Following hemisection and removal of the cerebellum to enable optimal access of solutions to MEC, brains were either cryoprotected in 30% sucrose for at least 24 hours until they had sunk, on in some cases, brains were stored in 0.02% sodium azide in PBS for up to 3 weeks before cutting. The use of sodium azide prevents loss of GFP fluorescence (Dumitriu et al., 2011). Sagittal sections were cut to 40µm using a freezing microtome. The processing of these sections depended on the fluorescence measure used to assess Arc expression: (1) direct GFP fluorescence, (2)

dual immunolabelling of GFP with a neuronal marker, and (3) dual labelling of Arc protein with a cell marker. Sections used for direct GFP fluorescence were mounted within hours in Vectamount and imaged within 24 hours to prevent loss of signal (Fig. 5.4(ix)). Sections for immunofluorescence were stored at 4°in phosphate-buffered saline (PBS) with 0.05% sodium azide.

5.3.4.1 *Choice of sections*

Appropriate sections for analysis were chosen based on identification of specific anatomical markers in the sagittal plane and on the quality of the tissue. Sections were typically taken from the left hemisphere, except in cases where tissue damage had occurred during freezing or cutting. All sagittal sections were mounted such that the pial surface of MEC layer I was approximately vertical. In this way, images from dorso-ventral regions could be acquired at equally spaced intervals using simple movement of the microscope stage. Sections for analysing visual cortex and MEC were chosen at the most lateral position at which the dentate gyrus granule layer is visible (Fig 5.5(a)). At this level MEC layer V extends the full dorso-ventral extent of the hippocampus whereas more medially it is truncated at the ventral extent. Sections containing dorsal CA1 were chosen in a plane in which the dorsolateral geniculate nucleus (LGN) is visible, because this is an easily identifiable landmark and CA1 is reasonably long and straight at this point (Fig 5.5(b)). Sections containing the dentate gyrus (DG) were chosen at the point where the lateral ventricle (LV) is elongated in the rostro-caudal plane (Fig 5.5(c)).

5.3.4.2 *GFP Immunofluorescence*

For GFP immunofluorescence I combined simultaneous labelling of GFP, using a primary antibody raised in rabbit, with labelling of NeuN, a neuron-specific protein (Mullen et al., 1992), using a primary antibody raised in mouse. Sections were incubated in PBST (0.4% Tween-20) for 15 minutes before incubation with 10% normal goat serum (NGS), 1X Roche blocking reagent in PBST. They were incubated for 3.5 hours in 1% NGS, PBST with Millipore mouse monoclonal antibody for NeuN (1:500) (A70 clone) and rabbit polyclonal antibody against GFP (1:2000) (Abcam 6556) at RT. Sections were washed for 3 x 5-10 minutes, then incubated in Goat anti-mouse IgG F(ab')$_2$ fragment, Alexa Fluor ©647 (1:400)(Invitrogen) and Goat anti-rabbit Alexa Fluor ©546 (1:400) (Life Sciences) for 2 hours at RT, then washed for 3 x 20 mins. Conjugates of F(ab')$_2$ fragment antibodies provide the advantage of reduced interactions with membrane containing Fc receptors (Invitrogen). Sections were mounted on slides with Vectamount and sealed with nail polish. Control sections underwent an identical procedure but with the GFP primary antibody removed (NeuN antibody still present). Following every immunostaining procedure, I checked tissue for the presence of non-specific staining that could result either from permeabilisation steps or secondary antibody incubation steps. See Appendix A for full protocols.

5.3.4.3 *Arc immunofluorescence*

For Arc immunofluorescence I used a monoclonal mouse antibody raised against the first 300 amino acids of the mouse Arc protein (C7, Santa Cruz Biotechnology). This type of antibody is highly specific to mouse Arc protein on mouse tissue, but detection can cause non-specific immunoglobulin binding, particularly of blood vessel walls, and high background staining. I tested two different methods to minimise these difficulties: (1) using a fluorophore-conjugated goat anti-mouse secondary approach, and (2) using the Vectamount MouseOnMouse (MOM) biotinylated antibody kit, which specifically reduces background staining caused by the use of mouse antibodies on mouse tissue.

In both approaches, sections underwent an antigen retrieval process in which they were incubated in sodium citrate buffer (pH6) at 80 degrees for 30 minutes. This served to increase access to the Arc antigen and reduce fluorescence caused by the secondary anti-mouse antibodies binding to immunoglobulins in blood vessels that weren't sufficiently cleared by perfusion. Following antigen retrieval, sections were blocked in 10% NGS on a shaker for 1 hour at RT, then incubated with the Arc antibody in PBST overnight on a shaker at 4°. Where the kit was used, the tissue was first incubated in a protein diluent prior to incubation with the Arc antibody. Following 3 x 5-10 minute washes in PBST, sections were incubated in 1:100 Alexa Fluor ©647 for 2 hours at RT then washed 2 x 10 mins in PBS. For the kit, sections were incubated for 10 minutes with 1:100 biotinylated antibody, which binds to the primary, followed by 2 washes then 5 minutes in Streptavidin-549 conjugate, which forms a streptavidin-biotin complex with the biotinylated antibody, followed by further washes.

The procedures both had advantages and disadvantages. Although a stronger signal: background ratio could be achieved with the goat anti-mouse fluorophore conjugate, there were occasionally cases of blood vessel staining. However, since the background non-specific staining could not be effectively reduced using the MOM kit biotinylated antibody, I made use of the first option, and subsequently did not analyse regions in which significant labelling of blood vessels occurred. As such, fewer mice were used for examining Arc protein expression than GFP protein expression.

In order to co-label cells for the purpose of identifying regions and Arc^{+ve} cells, I used a fluorescent Nissl stain, NeuroTrace, instead of the mouse NeuN primary antibody used with the GFP antibody. This served to avoid the possibility of cross-reactivity between the two primary mouse antibodies during the secondary antibody detection stage, and also reduced the possibility of damaging more vulnerable tissue. Sections were incubated in 1:100 NeuroTrace for 30 minutes then washed for 2 x 5 mins in PBS, followed by 1 hour in PBS.

5.3.5 *Acquisition of Confocal Images for Quantification of Arc/GFP expression*

To achieve an accurate and detailed assessment of GFP expression, I used a confocal imaging system that acquired three dimensional stack images of each specimen. All images were captured on a Zeiss LSM510 confocal microscope with a Zeiss 20x 0.8 aperture Plan-Apochromat objective (working distance 0.55mm) using a line scan that averaged over 4 lines to minimise noise resulting from the high gain necessary to pick up signal. The pinhole was set to 1 Airy

unit to minimise the collection of out-of-focus light. Image stacks were acquired at the optimal interval for the wavelength of interest, which was 0.89 µm for green wavelength light (for imaging GFP). For each image, 10-15 slices were captured, depending on the thickness of the tissue and penetration of the tissue by the antibodies. Fluorescence intensity considerably reduced beyond this. Sections were always exposed to the laser for similar amounts of time to ensure that any photobleaching affected specimens evenly. No photobleaching was observed throughout the experiment.

For each mouse and each region of interest (ROI), images were acquired of (1) direct GFP fluorescence, and (2) dual immuno staining of neurons, using NeuN, and GFP, using an anti-GFP antibody. For a subset of mice and ROIs, images were acquired of (3) dual staining of Arc protein and NeuroTrace. The fluorophores were excited using (1) an Argon 488 laser, (2) a HeNe 543 (GFP antibody) and HeNe 643 (NeuN) laser or (3) a HeNe 543 (Arc) and Argon 413 (NeuroTrace) laser.

Prior to imaging, the flatness of the specimen was always checked to prevent uneven illumination. To obtain optimal images for quantification, the gain and offset of the confocal system were set carefully to ensure maximum dynamic range. Because I ultimately wanted to compare images across baseline and non-baseline conditions, and across GFP^+/Arc^+ and GFP^+/GFP^+ animals, which express different levels of GFP at baseline, I used the settings most optimal for tissue with the brightest staining across multiple experiments and conditions. For all types of imaging, the region including visual cortex layer VI was chosen. One negative consequence of this was that since baseline levels of expression in GFP^+/Arc^+ mouse tissue are quite low, the dynamic range for the analysis presented in this chapter was relatively small. Where possible, I obtained all images for mice in a particular group of the experiment in the same microscope session to minimise effects of variation in the laser output. All images from a given section were always acquired in the same session.

5.3.5.1 *Defining anatomical boundaries for image acquisition*

Image stacks at 1024 x 1024 (460 x 460 µm) resolution were acquired at 9 locations (see Fig.5.6(a) (orange outlines)): (1) Patches (putative parasubicular area), (2-5) MEC layer II at ~ 400 µm apart along the dorso-ventral extent (such that images could potentially be stitched together), (6) MEC dorsal layer V/VI, (7) Cortex (visual) layer VI, (8) and layer II/III/IV, (9) CA1 (Fig.5.6(b)). For MEC layer II images, the microscope image was aligned such that the pial surface was just in view and oriented vertically. For each mouse tissue section containing MEC, 4 confocal images of 460 μm^2 were taken from the dorsal extent where the MEC borders parasubiculum down to the ventral extent. For each progressively more ventral image the microscope stage was moved using the LSM software such that the section moved 400 µm in the vertical direction. The image was then rotated and adjusted in the x direction to satisfy the above criteria. The patches (Fig. 5.6(a)(v)) were distinguished from layer II of MEC by the clustering of neurons and proximity of neurons to the pial surface. The dentate gyrus was also imaged (Fig.5.6(c)(x)), but using widefield illumination and a CCD camera (see next section).

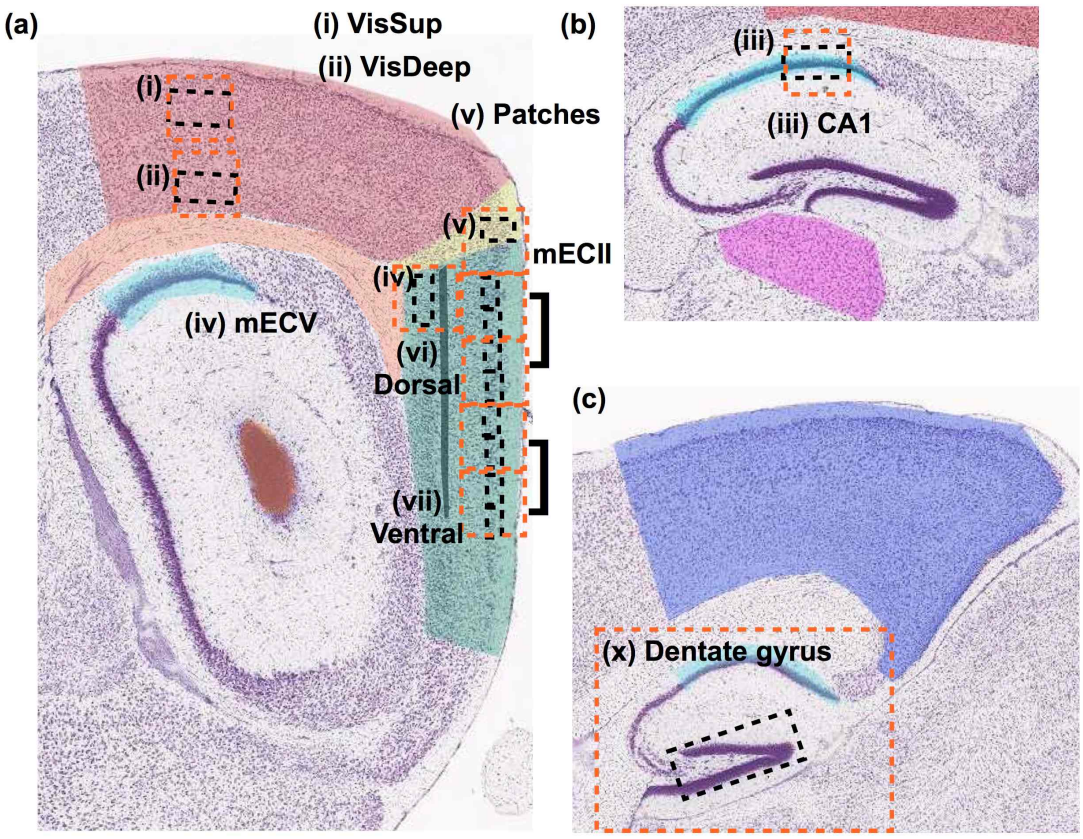


Figure 5.6: *Regions of interest for the quantification of gene expression* Images of Nissl-stained sagittal sections from the Allen Brain Atlas developmental reference atlas are overlaid by orange squares to indicate confocal 20x images acquired, and black rectangles to indicate the specific regions within them that were analysed. Each ROI was identified by certain features: (i) The superficial layers (II,III & IV) of visual cortex: the density and size of NeuN-labelled cells, (ii) The deep layers VI a and b of visual cortex: the density of cells and proximity to the alveus. (iii) The CA1 pyramidal layer: the presence of a thin layer of tightly packed cells. The most dorsal region was selected for imaging. (iv) MEC layer V/VI: delineated rostrally by a border with the fiber tracts (orange) and caudally by the layer IV laminar dissecans (dark green). (v) Putative parasubiculum: large patches of cells just dorsal to MEC. MEC layer II: medium to large darkly stained neurons (Insausti et al., 1997) (vi) dorsal: the most dorsal 600 µm (vii) ventral: the remaining layer II region (x) The dentate gyrus: thick layer of tightly packed cells. Whereas all other regions were imaged at 20x using confocal illumination, the larger dentate gyrus region was imaged at 4x magnification using widefield illumination.

5.3.5.2 *Acquisition of low-magnification images for visualisation and quantification of expression in the dentate gyrus*

Whole sections were also imaged at a lower 4x magnification using an Olympus microscope, Q-imaging colour fluorescence camera, widefield Xenon illumination and a Chroma TRITC filter. Low-resolution images were only used for quantification in the dentate gyrus where cellular expression is sufficiently sparse that cell detection can be performed with minimal axial resolution. As such, the analysis procedure differs for the DG compared with other regions. For visualisation purposes, images were stitched together using a combination of software tools including the ImageJ mosaic tool *MosaicJ* and grid stitch for automatic stitching. The stitched image was recombined using a linear blend. The image processing software,

GIMP, was used for manual alignment where automatic tools failed due to too small an overlap between images.

5.3.6 *Image Analysis*

To minimise experimenter bias and reduce the workload of processing a large number of images, once images had been cropped to regions of interest (Fig. 5.6 black boxes), all images were processed automatically and identically. The following sections will outline the image processing techniques used to quantify GFP and Arc protein expression from confocal microscope-acquired images (See Fig. 5.7 for an overview). These image processing techniques had two major purposes: (1) to correct for imperfections in image acquisition and to minimise the effects of intrinsic biological variation, slight differences in fixation and background noise, and (2) to enable the segmentation of the images to automatically detect cells. For GFP antibody labelling, the co-labelled NeuN could be used to detect neurons. For direct GFP, only GFP fluorescence could be used for cell detection. For Arc protein expression, I had the choice of using NeuroTrace or direct Arc protein expression for cell detection.

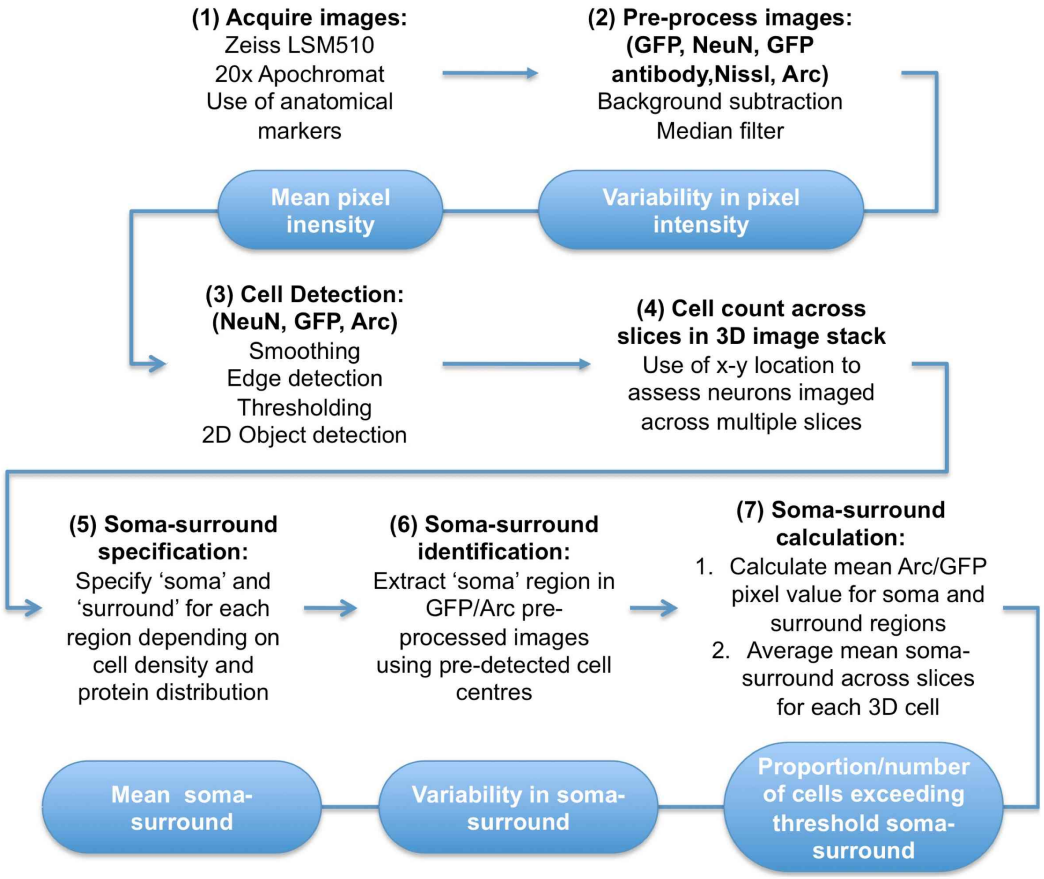


Figure 5.7: *Processing steps involved in quantifying protein expression in confocal images.* Blue ovals show the different measures available following each stage of processing for assessing absolute and relative protein expression across regions. ROI-wide measures are available following stage (2), and soma-surround measures, which are invariant to overall image brightness, following stage (7). Cell detection processing stages were used only on pre-processed direct GFP, NeuN and Arc images, with the results applied to detect GFP and Arc protein expression. The calculation of thresholds was based on comparing antibody staining to controls to give signal to noise estimates and will be described in Section 5.3.7.2.

5.3.6.1 *Procedures for image handling*

All confocal images were split into individual channels and stored as 8-bit grayscale tiff files. All image analysis was performed using ImageJ plugins (Background subtraction, median filtering, FeatureJ, Pairwise Stitching, Automatic thresholding), the Python Numpy library or the Python Image Library.

5.3.6.2 *Specifying regions of interest for quantification*

To extract identically sized areas of interest for each ROI so that images could be analysed equivalently, all images were rotated, where necessary, and cropped post-acquisition (Fig. 5.6 black areas). To prevent bias, where possible I selected regions for analysis using cell-related rather than Arc-related staining. For MEC layer II I selected 6-8 regions (depending on tissue integrity and curvature) of size 400 x 200 pixels (~ 90 x 180 µm) along the dorso-ventral axis. In native GFP images the location of layer II was estimated based on proximity

to the pial surface. In NeuN co-stained images, layer II was selected based on the presence of large, dense heterogenous cells (see Fig. 5.6 (v-vi)). In some cases, very few large cells were present at a particular dorso-ventral extent possibly due to a thinning of the layer. For MEC layer V/VI, images were cropped to 550 x 200 pixels (~90 x 250 μm) to include the most dorsal and central region. As such, layer V/VI images typically do not include putative layer Va or layer VIb expression, but this was not formally defined. In direct GFP images, proximity to the hippocampus and postrhinal border were used to determine position. For GFP antibody images, NeuN stain clearly distinguished homogenous cells in layer III from the smaller, denser cells in layer V/VI in co-stained images. For CA1 the pyramidal layer could be identified in both direct and NeuN stained images and was cropped to 800 x 250 pixels (~180x360 μm). In visual cortex a region of 900x600 pixels (~405 x 270 μm) was selected to include layers II/III/IV, which could be identified by baseline cellular fluorescence in direct GFP images and by cytoarchitecture in NeuN staining. In layer VI a region of 800 x 500 pixels (~360x225 μm) was selected such that the bottom edge aligned with the border of the corpus callosum.

I examined Arc protein expression in a subset of regions considered for GFP expression: MEC layer II (dorsal and ventral) - both 1000 x 200 pixels (~90x450 μm), MEC layer V/VI (~90x450 μm) and Visual layer VI (900 x 600 pixels, ~405x270 μm). NeuroTrace could be used to determine the correct location.

For each region of interest image stacks sufficient to record all in-focus areas of the tissue were acquired in the axial (z) plane. To equate stacks across mice and across regions, the highest mean intensity image slice within the stack was defined as the centre and where possible, the 4 slices each side were also chosen to create a new image stack of thickness ~10 μm.

5.3.6.3 *Image pre-processing to minimise variability due to noise*

Noise can result from inevitable statistical fluctuations, limitations in resolution leading to increased diffraction of light, inhomogenous contrast due to non-perfect distribution of dye, and background gradients due to non-uniform illumination (Meijering, 2010). An example of the effect of noise on the histogram of an image containing 3 artificial cells with just 3 differing intensity values is shown in Fig. 5.8(i-ii). The aim of pre-processing is to restore as much of this information as possible.

Median filtering was applied to the images to reduce impulsive background noise whilst preserving edges (Lim, 1990). Median filtering is a nonlinear digital filtering technique in which pixels are replaced by the median of their surrounding pixels within a user-specified radius. As shown in Figs. 5.8 and 5.9(iii), median filtering enhances the detection of multiple peaks in an image histogram that correspond to signal information. This is particularly useful in this context because the aim is to identify small but supra-pixel sized regions of Arc or GFP protein expression and to obtain accurate estimates of the level of this expression.

Background subtraction was applied to images to remove variation in background illumination from the image and to maximise the sensitivity of analysis tools to differentiate signal from noise. This provided a clear enhancement of the cells relative to their background (Figs. 5.8(iv) and 5.9(iii)). The ImageJ 'Subtract Background' plugin tool (Schindelin et al., 2012)

uses a rolling ball algorithm based on that described by (Sternberg, 1983) with a user-defined radius to determine areas of background to be removed. To determine an appropriate radius, the largest cells found in each area were identified and measured to give a value between 30 and 50 pixels (13-22 μm). These techniques could be used successfully on all image types and aided in visualising the co-labelling used to detect GFP and Arc protein expression within cells (Fig. 5.9(iv)).

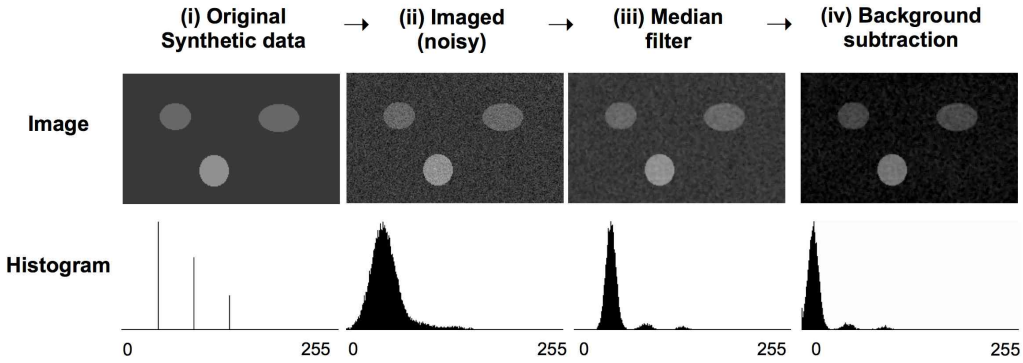


Figure 5.8: *Pre-processing steps to minimise variability in background noise across images and to enhance signal detection.* (i) A synthetic image of cells (upper) is shown alongside its histogram (lower), which shows the distribution of pixels across pixel intensities, revealing the main features of the image. (ii) Artificial noise is added, to reflect the imaging process, which typically makes feature extraction more difficult. (iii) Median filtering restores the peaks in the histogram corresponding to the lighter circles. (iv) Background subtraction removes non-signal uniform background noise from the image.

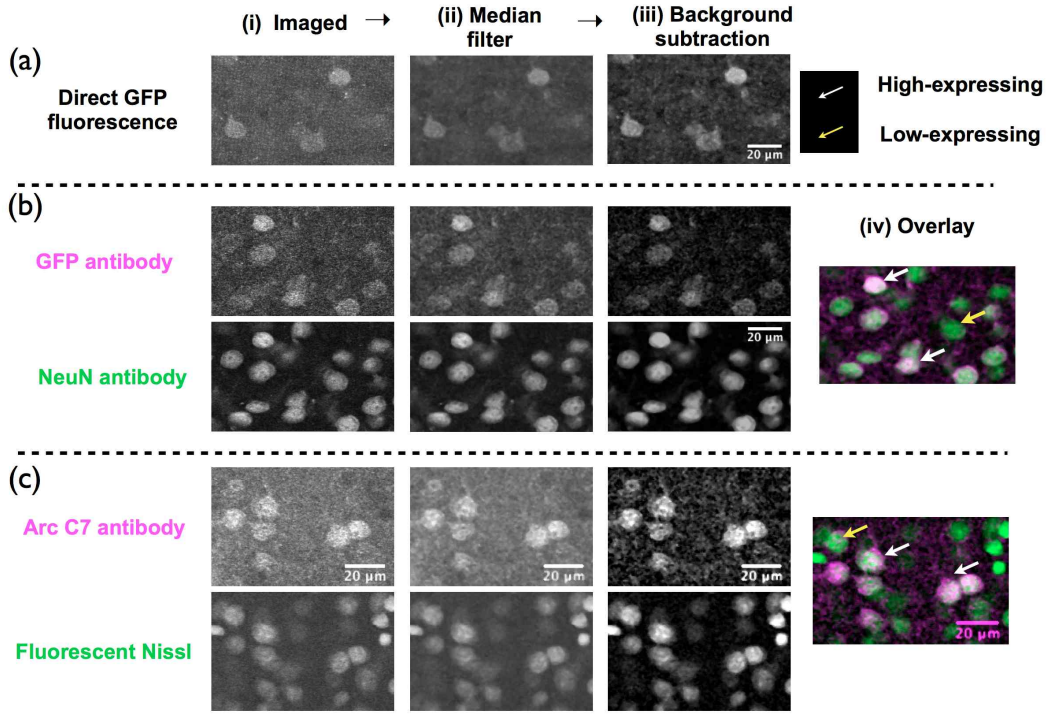


Figure 5.9: *Pre-processing steps applied to confocal images of GFP and Arc protein expression.* Example confocal 20x single slice images showing (a) direct GFP fluorescence, (b) GFP expression using an antibody against GFP and NeuN-specific antibody to label neurons and (c) Arc protein expression using mouse monoclonal Arc C7 and NeuroTrace fluorescent Nissl stain. (i) Original images, (ii) following median-filtering and (iii) following background-subtraction to enhance signal detection. (iv) Cell staining (green) and GFP (b) or Arc (c) antibody (magenta) images have been overlaid to show the cellular distribution of protein expression. White arrows emphasise examples of neurons or cells that express high levels of Arc or GFP, whereas yellow arrows indicate examples of low-expressing cells. Scale bar 20µm. All images shown were adjusted for brightness and contrast for this figure.

For some of the sections labelled with a mouse antibody for Arc immunohistochemistry, blood vessels were also labelled by the goat anti-mouse secondary antibody. A semi-automated procedure was developed to detect and remove these blood vessels such that areas of the image that were affected were not included in subsequent quantification (see Fig. 5.10). As a result, pixel intensity calculations within automatically detected cells in the Arc image only include Arc antibody labelling, and not background labelling of tissue by the Goat anti-mouse secondary antibody.

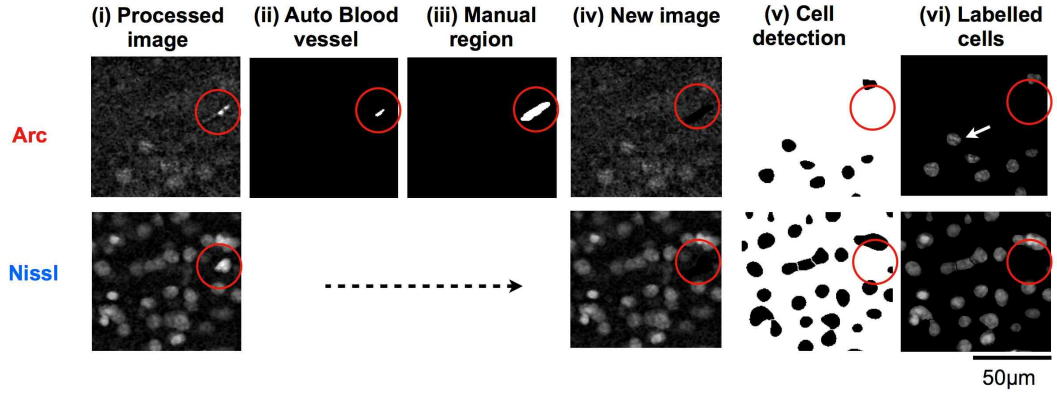


Figure 5.10: *The procedure for blood vessel removal from Arc protein expression images.* (i) Processed confocal single slice images (after background subtraction and median filtering) of Arc protein expression and Nissl staining. (ii) Images showing automatically detected high intensity tube-like structures from images in (i), outlined by a red circle. (iii) Images showing manually adjusted blood vessel region. (iv) Adjusted images from (i) in which detected blood vessel region in (iii) has been removed using an image mask. (v) Cell detection mask image, which does not include any cells in the region of the labelled blood vessel. (vi) Image showing Arc and Nissl labelling at the location of detected cells. No pixels associated with the detected blood vessel fall within detected cell regions.

5.3.6.4 *Detection of cells for cell-based protein expression analysis*

Following pre-processing, three cell detection methods were available to enable the analysis of somatic Arc and GFP expression in the confocal stack images; (1) manual detection, (2) automated volumetric detection, and (3) 2D object detection, followed by identification of unique cells across images in a stack using the relative location of their cell centres across slices. Measures (2) and (3) can be compared with a manual count of neurons detected using NeuN to assess the most sensitive and specific tool (see next section).

1. Manual detection involves a blind experimenter marking individual cells as they proceed through a confocal stack, ensuring that each cell is only marked once. Although this method is likely to be consistent across small datasets, it is subjective, and would be very time-consuming for this scale of image analysis. Manual detection was therefore only used on a small scale to validate the results of automated detection.

2. Automated 3D volumetric cell detection, which can be achieved using ImageJ's 3D Object Counter, involves the detection of cells across slices (in z) that occupy a similar location in the x-y plane. This tool enables the user to specify an appropriate volume for cells that it detects such that results can be limited to the expected size range, but it does not differentiate between the area covered in x-y and the number of slices in z.

3. 2D object detection: Given the limitation of 3D detection, I decided to split the process of cell detection into a 2D object detection stage, followed by use of a custom-written tool to detect similar cell locations across images within a stack and assign these to the same cell.

To quantify native GFP, cells were automatically detected from 2D images using GFP expression alone, thus limiting detection to high-GFP expressing cells. NeuN, a neuronal

marker, was used for the GFP immunofluorescence methods, such that both high- and low-expressing neurons could be analysed. For Arc images, I trialled detecting cells using the NeuroTrace images, but found in MEC layer II that many non-neuronal cells are labelled and that labelling of NeuroTrace does not provide an accurate border for measuring protein expression. I therefore detected Arc^{+ve} cells using a semi-manual, semi-automated approach from the Arc protein expression images (described in the next section).

2D object detection: Effective cell detection in two dimensional images is necessary for both ImageJ 2D and 3D object detection tools and requires several steps. Cell detection tools typically threshold an image to distinguish an object from background. This can be problematic when quantifying protein expression because objects are not uniformly different from background. NeuN, GFP and Arc are not expressed uniformly within the cell body, so first, a smoothing step was used to make the cell body more uniform prior to edge detection (Fig. 5.11(i)). The radius over which to smooth each image varied across different measures and different regions but was between 6-10 pixels. For this step, images were processed as stacks using two-dimensional filters.

Following smoothing, an edge detection method was used to enhance the outline of cells prior to thresholding (Fig. 5.11(ii)). To identify cell outlines, the Laplacian of each image was calculated using an ImageJ 'FeatureJ' plugin written by Erik Meijering. This measures the 2nd spatial derivative, and therefore highlights regions of high intensity change such as edges. The resulting image was thresholded to only extract objects defined by strong edges (Fig. 5.11(iii)).

A limitation of the confocal images acquired, as well as the need to threshold them, is that the resolution was limited by the 0.8 aperture of the objective. The effect is that nearby cells can appear as the same cell. Following thresholding, images were processed through a watershed filter to separate, neighbouring cells into distinct objects (Fig. 5.11(iii)), where possible.

For 2D cell detection, the resulting image was processed by the ImageJ particle analysis tool, which detects cells and provides details of their location and size. This tool can be used to only extract cells within a given size range. Values for appropriate cell size in MEC were determined based on the detailed research of Gatome et al. (2010), who carefully calculated and analysed the statistics of the different types of cells present, and by plotting the distribution of thresholded cell sizes across different regions. Cells smaller than a manually set threshold (Fig. 5.11(iv) (red vertical line for thresholds) were not included in the subsequent particle analysis. The tool produced final mask images only showing cells satisfying size criteria (Fig. 5.11(v)). All image slices within stacks were processed in this way (see Fig. 5.12(A))

I attempted using this automated approach for Arc protein expression. However, Arc protein labelling in low-expressing tissue is often limited to a ring surrounding the somatic region (the soma could be identified using NeuroTrace), which was difficult for the detection algorithm to detect. To augment the analysis, each image was visually scanned for the presence of undetected but viable Arc^{+ve} cells, which were then marked for subsequent analysis. Because I used fewer slices (7 vs. 9)within each stack for Arc images due to poor antibody

penetration, I analysed protein expression at the same locations in each slice throughout the stack and detected the validity of the detected cell volume at a later stage.

3D object detection: For 3D cell detection, the resulting 2D image masks (Fig.5.11(v)) were input to the ImageJ 3D object counter. Appropriate cell volumes to exclude non-cells were calculated based on cell radius ($Volume = 4/3\pi r^3$), as estimated from approximate cell size used for 2D detection.

Identifying cells across slices: A limitation of 2D cell detection compared to 3D cell detection when using 3D image stacks is that 2D detection does not take into account the fact that most cells exist across multiple images, and so should only be analysed as one cell, not multiple. The ImageJ object detection plugin identifies and reports the coordinates of all cell centres detected. I developed a tool using Python to sort through all cell centres for each image stack and to identify all cell centres within a certain range of one another across slices. Cells with similar centre locations were subsequently analysed as single cells (see Fig. 5.12(B-C) coloured arrows and their respective coordinates) and any measures drawn from them were averaged across the slices in which they were found. For Arc images, cellular regions across all slices were automatically analysed as one cell unless the soma-surround score for a region was less than zero.

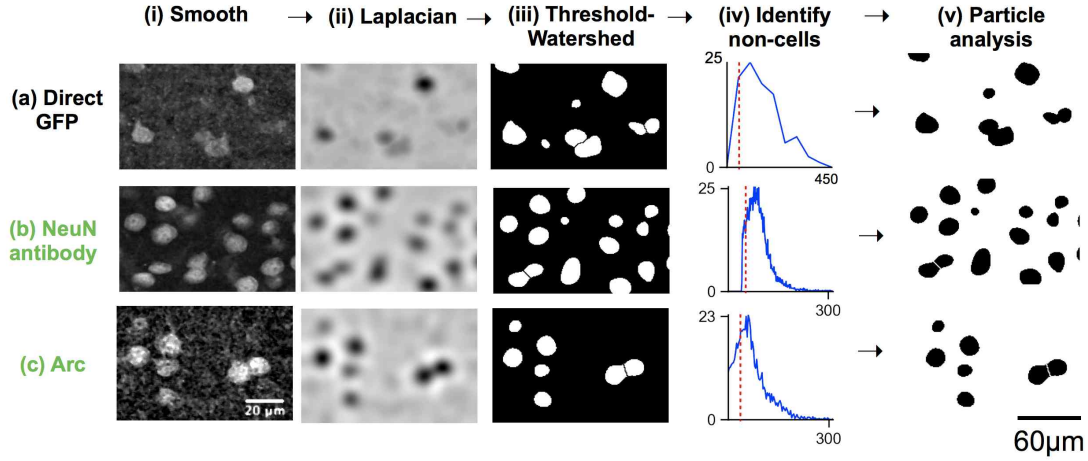


Figure 5.11: *Processing steps for cell detection.* Images of (a) direct GFP, (b) NeuN antibody and (c) Arc labelling show the results of sequentially applying ImageJ plugins that perform the following processes: (i) smoothing with a radius of 6-10 pixels (depending on cell size and density), (ii) edge detection using a Laplacian filter (FeatureJ), and (iii) automatic thresholding using the median to estimate the threshold, followed by watershedding to segregate overlapping objects for which the boundary could not be resolved. (iv) Histograms show the distribution of cell size (in area μm^2) of cells detected from (iii) using the ImageJ particle analysis. The manually set thresholds are shown by the red vertical line. (v) Images show detected cells with sizes exceeding the set threshold in (iv).

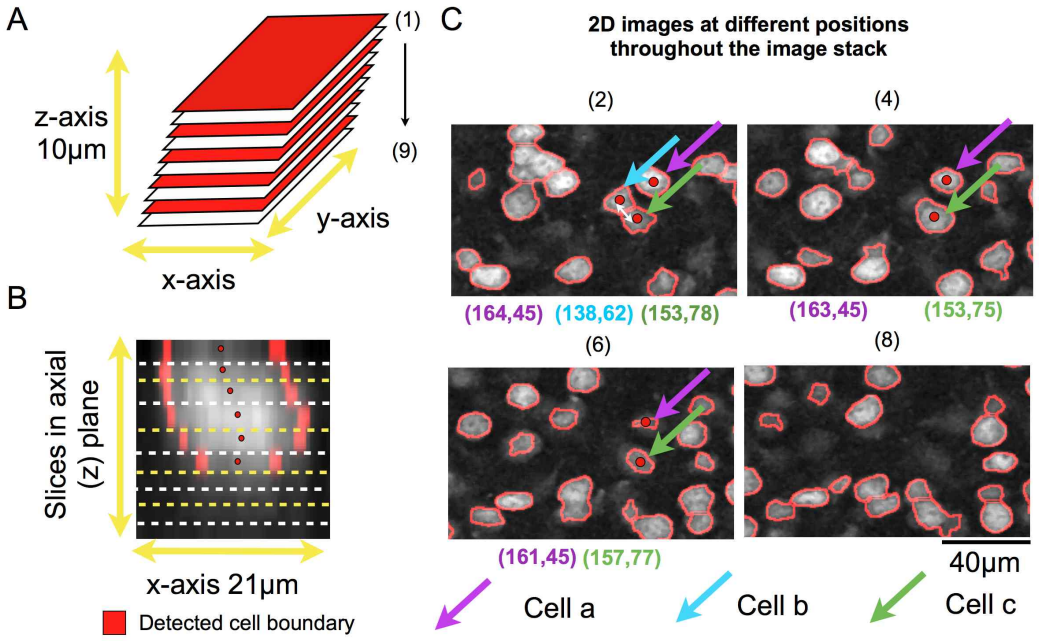


Figure 5.12: *Identification of neurons across slices within stacks.* (A) Schematic showing the dimensions of image slices in a confocal stack, which was approximately 10 µm thick (z-axis). (B) Image in the x-z plane showing a single cell and its detected boundaries throughout different image slices of a confocal stack. Red dots mark the identified centre in each image. Data for each of the cells in these stack slices was averaged across slices. (C) The x-y slices (2,4,6,8) from (A) are shown overlaid with cell boundary outlines. Three cells that are present across slices and their respective coordinates are shown. Cells within a Euclidean distance of 5 pixels are considered as the same cell.

Validating cell detection methods: The success of the two automated cell counts (3D vs 2D plus merge across slices) was assessed by comparing the results with those of a manual cell count. Fig. 5.13(A) shows the outlines of all neurons detected using NeuN, overlaid on the original images for each of the different methods. The 3D volume count using ImageJ plugin identified very few true cells, possibly because neighbouring cells are very near to each other, which renders segmentation difficult (Fig. 5.13(B)). The 2D count, followed by identifying cells in similar x,y locations throughout the axial extent as a single cell, detected the majority of cells detected manually. The only failures resulted from faint staining and close proximity to other cells, as well as an example of oversegmentation. It also detected cells at the edges of the image that were not included in the manual count. This analysis did not take into account that cells could be misidentified in some slices but correctly segregated from others in other slices of a stack. Since the majority of the analysis relies on the detection of cells compared to their surround, this error probably has little effect on the overall outcome.

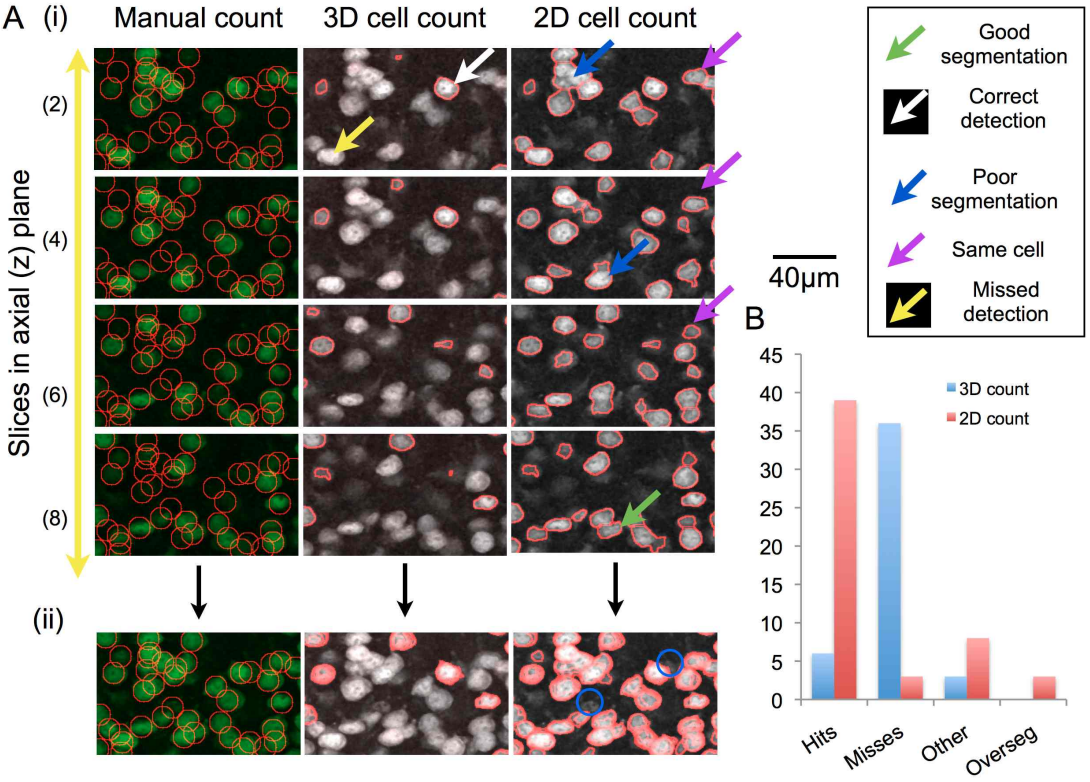


Figure 5.13: *Validating cell detection methods using NeuN labelled images.* (A)(i) Images show NeuN labelling throughout 4 alternate slices in a confocal stack. For manually counted 'ground truth' images (left), red outlines show the location of a cell that is present in at least one of the slices. 42 unique cells were identified. For the 3D cell count (centre) and 2D cell count, red outlines show cells detected in that particular plane. (ii) For all the counts, the images have been projected into a single plane using a maximum intensity projection to show all cells detected in the stack. Blue circles mark failures on the 2D image count, which detected almost all of the manually detected images. (Cells on edges not included). Purple arrows show examples of cells in 2D that would be counted as the same cell in 3D. Yellow arrows show cells that were not detected by the algorithm, whereas white arrows show correct detection. Blue arrows show cells that have not been correctly segmented, whereas green arrows show correctly segmented cells. (B) Bar chart showing the number of cells correctly detected (hits), missed, detected that were not manually counted (Other) and falsely detected due to oversegmentation (Overseg) is shown for the 3D cell count (blue) and 2D cell count (red).

5.3.6.5 *Defining a soma-surround region for quantification*

The simplest way to quantify a fluorescence image is to consider the distribution of pixels across an ROI and infer protein expression based on mean values. Cell-based measures of protein expression are important as a complement to ROI-wide pixel intensity measures because they provide insight into density of expression and cell type and are not influenced by cell density. However, cell-based measurements that depend purely on absolute pixel intensity are still limited by possible differences in laser power and immunofluorescence efficacy, even following image pre-processing steps. As such, I employed a simple but more sensitive method that I adapted from the cell detection analysis used by Wang et al. (2006). This required identification of a 'surround' area encompassing pixels within a limited distance of the soma outline (Fig. 5.14(A)). For each fluorescence measure, three images were necessary

for this step: the preprocessed Arc or GFP image (for measuring soma-surround expression), the preprocessed NeuN image, for identifying 'soma' and 'surround' regions for a given cell centre, and the cell detection thresholded NeuN, direct GFP and Arc images, for providing the regions or central coordinates of detected cells. The method for identifying the surround region (and variants), which I implemented in a tool written in Python, is outlined below.

1. Determine appropriate 'soma' region. There were two options that were chosen depending on how closely detected cell outlines matched actual cell outlines.

    a) Use of the 'detected' cell image (see Fig 5.14 (B)(i)) for an example). This method assumes that the detected region of the cell is an appropriate approximation of the true cell region. It is appropriate for the majority of NeuN images, where the average original labelling closely matches the average detected cell borders (see Fig. 5.14 (C)) However, for direct GFP, although in the example shown in Fig 5.14 (B)(ii), there is a very close correspondence between the detected image and the GFP^+ cell, this was not the case on average (see Fig. 5.14 (C)). As such option (b) was used instead.

    b) Use of an estimated cell region. This method identifies a 'cell' region using the preprocessed (not thresholded) NeuN/original GFP/original Arc images and more closely follows the method of Wang et al. (2006). First, a potential cell is estimated by using the algorithm-detected (or manually detected for Arc) cell centre (see Fig 5.14 (B)) and the maximum size of a cell in the region (data acquired from histograms in Fig. 5.11(iv)). Second, all pixels within this potential cell that are greater than half the maximum pixel intensity in the NeuN/direct GFP/Arc cell detection image are classed as 'cell' pixels and together they comprise the 'cell'. 50% appears to be an appropriate value to capture the majority of pixels that are truly in the cell. The plot in Fig 5.14 (B) (iii) shows the pixel intensity at each point along the red dotted line (ii). The blue line marks the boundary between pixels identified as 'cell' pixels or not. If the number of pixels satisfying this criteria is less than the cell size criteria applied to cell detection (data acquired from histograms in Fig. 5.11 (iv)), then the cell is not included in the analysis. One possible problem with this is that if the maximum value is comparable to background intensity then most pixels will be included, so it is important to only detect true cells at the previous cell detection stage (direct GFP), or to identify 'cells' that this could apply to (Arc).

2. Determine appropriate 'surround' region. This depends on the 'soma' measure chosen and on ROI. The 'surround' is a circular region defined by a bounding box between 1.5 and 2 times (depending on the factors below) the size of the bounding box defined by the original detected cell. To be meaningful, it should not capture true cell pixels but only the immediate region surrounding the cell

    - Cell measure: As 'soma' regions were estimated differently for direct GFP/Arc and NeuN, a smaller 'surround' diameter was chosen for the former to take into account the fact that the cell region may already be an overestimate of cell size.

For NeuN, the 'surround' region did not include the pixels of the detected cell, a narrow region surrounding the detected cell, or pixels of other cells. For direct GFP/Arc, the surround region did not contain pixels of the 'soma' as defined by pixel intensity measures, or pixels of neighbouring cells.

- ROI: Density of cells also affects the most appropriate definition of the 'surround'. In the CA1 pyramidal layer, calculating a surround was not appropriate because cells are so densely packed that it would ultimately incorporate labelling from neighbouring cells that couldn't necessarily be segmented from background. Instead, the average extra-somatic pixel intensity was used as a 'surround' for CA1.

3. Calculate, for each detected cell in 2D images, a soma-surround measure. To acquire soma-surround measures for actual cells across multiple slices, in 3D, these measures were averaged across slices for each cell.
   a) Use ImageJ plugin/manual search to identify cell centres in 2D image slices from cell detection-processed images.
   b) For each defined cell calculate the mean pixel intensity of pixels in Arc/GFP image that correspond to pixels in the NeuN or detected cell image assigned to the 'soma'.
   c) Subtract from this the mean pixel intensity of pixels in Arc/GFP image that correspond to pixels assigned to the 'surround'. I used subtraction rather than the division used by Wang et al. (2006) because the surround region for low-expressing tissue was often near-zero, which would lead to incorrectly high soma-surround scores.
   d) This gives a measure, for each cell, that reflects how much more GFP or Arc is expressed in cells than in their surrounding non-cell region.
   e) This can be applied to all cells in a slice to highlight high and low-expressing cells (See Fig 5.14 (D)).

   The final stage is to average soma-surround measures across animals within a group. For direct GFP and Arc, at least 3 cells had to be detected per region for the mean and variability in soma-surround measure to be included in the final analysis for that mouse. If fewer than 3 cells were detected per region for a given mouse, the count was recorded as such, but the data for mean and variability were excluded from these calculations, to prevent anomalous data obscuring more consistent results.

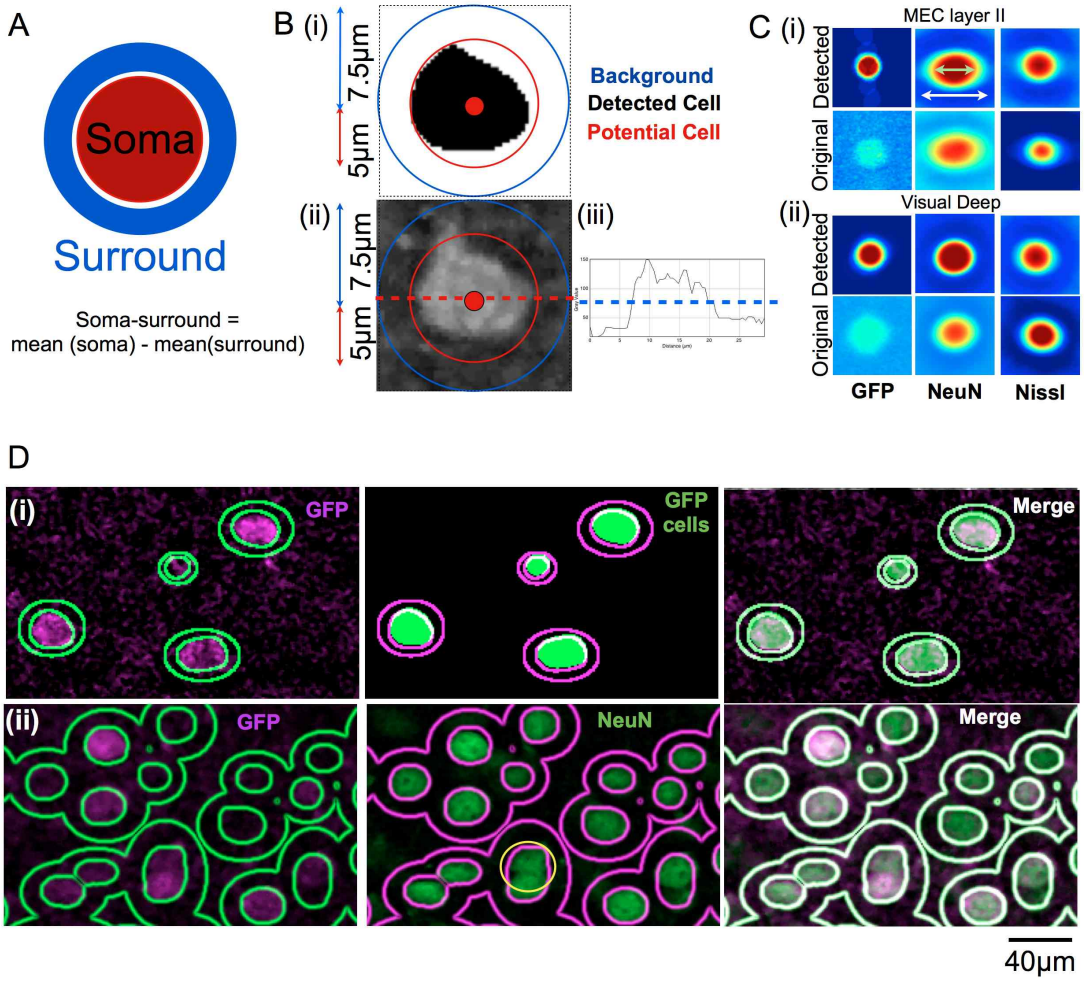


Figure 5.14: *Quantification using soma-surround method.* (A) Schematic showing 'soma' and 'surround' region allocations. (B) Example images showing (i) a detected cell (upper) and (ii) the original labelling (lower) for a direct GFP-imaged cell. Images are overlaid by the potential 'soma' region (red) and 'surround' region for which location depends on the cell centre (red dot) and size depends on the size and density of cells in that region.(iii) Cross-section plot showing the pixel intensities along the red dotted line in (ii). The blue dotted line shows the half-maximal pixel intensity, which can be used to distinguish soma from surround. (C) Images show the 60 x 60 pixel region, averaged across all detected cells, surrounding a detected cell centre (upper) and the original GFP (left), NeuN (centre), or Nissl (right) image that the cell was detected from (lower) for (i) MEC layer II cells and (ii) visual cortex deep layer cells. Arrows show average width of the 'soma' (green) and 'surround' (white) at 13.5 and 21.5 µm, respectively. Due to the semi-manual analysis of Arc images, these images could not be generated. Nissl images have been included instead to highlight the difference in detected cell shape between NeuN and Nissl-detected cells. (D) Overlays show cell boundaries and approximate chosen 'surround' regions for cells detected in (i) direct GFP and (ii) GFP antibody (magenta) (NeuN: green) images. Yellow circles mark cells not detected. Overlays could not be generated for Arc protein images because of the slight difference in analysis method.

5.3.7 *Quantifying region and cellular expression*

5.3.7.1 *Summary of measures*

Following image pre-processing, cell detection and soma-surround calculation, the characteristics of all defined 'cell's, including cell number, diameter, mean cross-sectional area, mean pixel intensity, soma-surround measure, shape and density, were calculated. These measures have been used to quantify Arc and GFP throughout the chapter (See Fig. 5.7). Their advantages and disadvantages are shown here:

1. Mean pixel intensity across an ROI: this provides the most coarse measure of GFP/Arc expression. High levels typically reflect higher expression. However, the disadvantage of this measure is that uneven illumination across the cover slip can directly cause variability in this value, as can different intensity of non-signal background staining.

2. Variability in pixel intensity across an ROI: this provides an indication of the variability of expression within animals and the sensitivity of staining. Increased variability within animals suggests that there is a true signal and background.

3. Mean soma-surround for each mouse for each region of interest: this provides a measure of somatic expression and is a useful measure because it is minimally affected by differences in laser intensity and non-signal background variability in fluorescence, as well as changes across a section that could result due to uneven illumination. However, it is limited by the accuracy with which it can be defined, and by the presence of protein expression in proximal dendrites. It also provides no indication of protein expression in the neuropil.

4. Soma-surround standard deviation: this gives an indication of how variable Arc/GFP expression in this region is and indicates whether Arc/GFP is expressed at different levels across cells.

5. Proportion of cells exceeding low soma-surround threshold: this threshold provides an added measure that theoretically isolates cells actually expressing protein from those that do not. It was set by analysing the distribution of control tissue data and only allows scores in the top 5% of control values and greater.

6. Proportion of cells exceeding moderate and high soma-surround thresholds: these give an indication of the number of cells expressing high levels of Arc and are therefore comparable to thresholded cell counts used in other research using immuno-precipitation. The thresholds were set using the median and specific upper percentiles of the distribution of scores. They provide more information about the highest-expressing cells than the mean or standard deviation can provide.

5.3.7.2 *Validation of cell-based quantification measure*

These measures were compared by examining their ability to distinguish true protein expression from control (Fig. 5.15). Control sections containing visual cortex layer VI were imaged

and processed identically to normal immunofluorescence samples. The control used for the GFP antibody was its secondary-only control, for direct GFP was fluorescence from WT animals expressing no GFP, and for Arc was tissue from KO animals expressing no functional Arc protein (Wang et al., 2006). For Arc protein expression analysis, no cells were automatically detected. To achieve a soma-surround score for the purpose of comparison, protein expression was sampled from regions marked out by NeuroTrace-labelled cells.

The distribution of GFP protein expression when imaged directly differs considerably from its control on both ROI-wide and cell-based measures (Fig. 5.15(A)(a)&(B)(a)). ROI-wide mean fluorescence intensity (t_{20}, $p<0.0001$) and mean soma-surround ratio (t_{20}, $p=0.002$) are both greater for direct fluorescence in Arc^{+} /GFP^{+} than in WT animals (Fig. 5.15(C)). The number of cells detected in the control condition was close to but not zero and likely resulted from small fluctuations in intensity across the WT tissue. The proportion of cells exceeding threshold was calculated by dividing the number exceeding threshold by the total number detected across both groups.

GFP antibody labelling is also significantly higher than secondary-only antibody labelling (Fig. 5.15)(A)(b)&(B)(b)). ROI-wide mean fluorescence intensity (t_{20}, $p=0.0016$), variability in fluorescence intensity (t_{20}, $p=0.02$), mean soma-surround ratio (t_{20}, $p=0.002$), variability in soma-surround ratio (t_{20}, $p=0.0006$) and proportion of neurons exceeding threshold (t_{20}, $p=0.03$) are all significantly greater for GFP antibody labelling (Fig. 5.15(C)). Of note, the soma-surround measure provides a direct means of testing for any interaction between NeuN and GFP antibody staining as the secondary-only control provides an indicator of non-specific binding of the secondary antibody, either to mouse tissue or to the NeuN primary. This measure should be zero if the staining is independent but it was slightly higher than zero, which suggests there may be a very small interaction.

The comparison between Arc protein expression and its control gives a slightly different result. True protein expression can be most clearly distinguished using mean ROI-wide pixel intensity measure ($p=0.07$, $n = (2,3)$), whereas the soma-surround measure does not indicate a clear difference ($p>0.05$) (Figs. 5.15(A)(c)&(B)(c), (C)). The proportion of cells exceeding threshold was calculated by dividing the number exceeding threshold by the total number detected across both groups. Given these data, it is likely that Arc protein is expressed at very low levels in visual cortex in Arc^{+} /GFP^{+}, and that when it is expressed, it is not necessarily in cell bodies but in the neuropil. The non-zero result for GFP^{+}/GFP^{+} mice is likely due to random fluctuations in non-specific binding of the secondary antibody to the tissue. It suggests that the analysis of these images could be optimised to eliminate any 'cells' from being detected in Arc KO tissue.

The relative distributions of GFP/Arc labelling and control were used to choose thresholds that distinguish 'low-expressing' cells from cells with no detectable expression exceeding chance level for the analysis throughout this chapter. Median and high thresholds were calculated by pooling all data from all animals across all regions and finding the 50% and 75% (for direct GFP and Arc) or 90% (GFP antibody) percentiles.

5.3.8 *Statistical methods*

For the majority of comparisons across regions, I used a one-way independent-measures ANOVA, followed by post-hoc Tukey HSD tests if the ANOVA p value was lower than α. I could have used repeated-measures ANOVA tests on the basis that data from different regions was collected in the same mouse. However, there is no clear evidence that the behaviour of Arc expression should always be related in the same way across regions. Where previous results and data from the literature provided clear predictions regarding differences, I used a one or two-way *a priori* t-test, as stated, on the relevant data. This included comparisons of protein expression with their control. All statistical tests for ANOVA and post-hoc t-tests or Tukey HSD tests were performed using R (R Development Core Team, 2008). An α value of 0.05 was used as a threshold for significance. All graphs are presented as mean $\pm$ standard error.

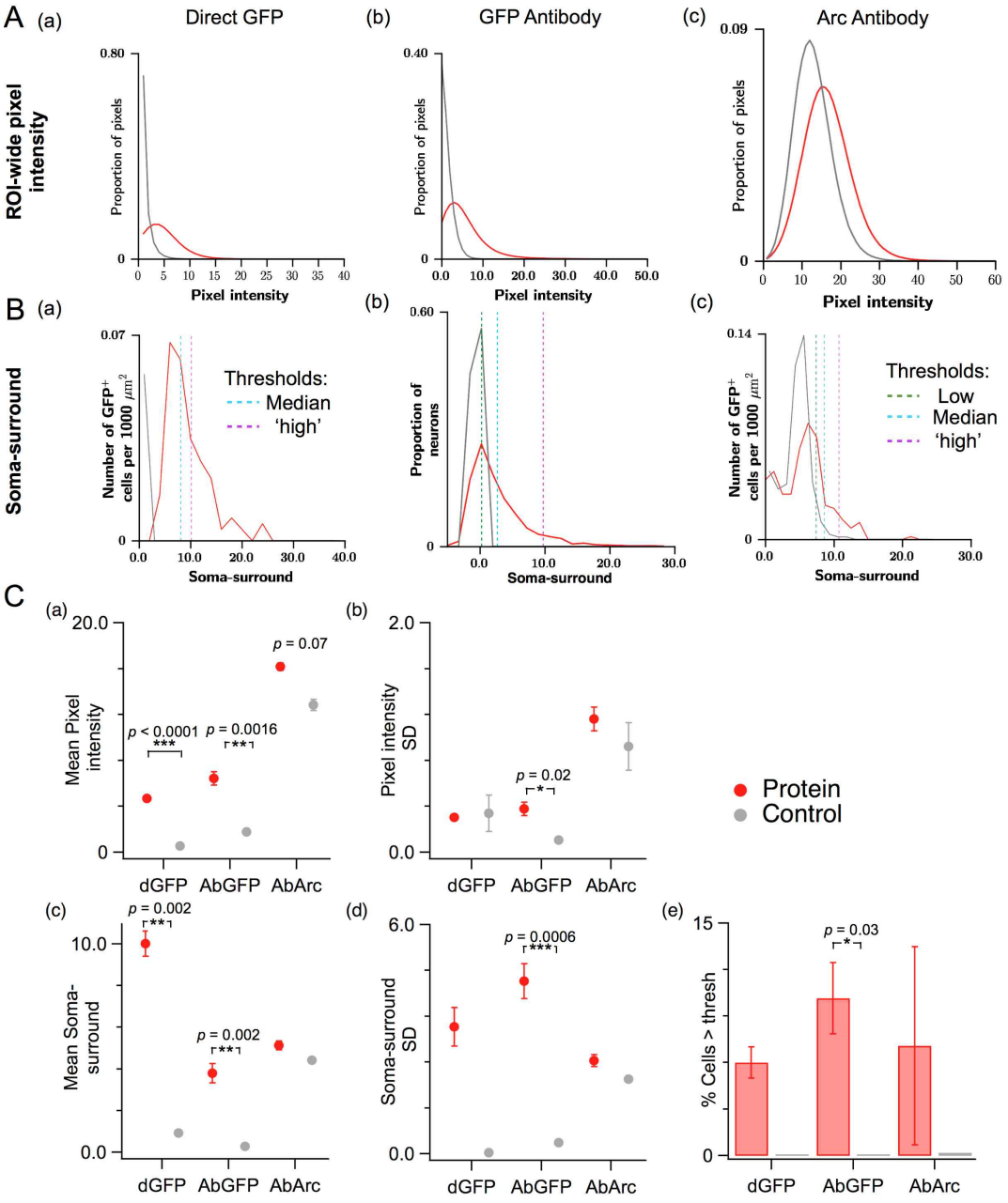


Figure 5.15: *How the measures compare when distinguishing true protein expression from control* (A-B) Histograms show the distribution, averaged across all animals, of measures comparing GFP or Arc protein expression (red) and its respective control (grey) in the deep layer of visual cortex. The following measures were used: (A) ROI-wide pixel intensity, (B) Soma-surround score. For direct GFP images (a), directly imaged GFP fluorescence from Arc$^+$ /GFP$^+$ mice (n=8) was compared with directly imaged fluorescence from WT tissue (n=3). For the GFP antibody (d-f), antibody staining from Arc$^+$ /GFP$^+$ mice (n=8) was compared with secondary only staining from a subset of the animals (n=3). For the Arc antibody (g-i), antibody labelling from Arc$^+$ /GFP$^+$ mice (n=3) was compared with that from GFP$^+$ /GFP$^+$ mice (n=2). The green vertical line represents the 'low' threshold, a value which almost no control pixels or cells exceed. The magenta vertical line represents the 'high' threshold, a value at which ~ 10% of the cells detected exceed (plotted in C). (C) Summary measures for each of the three fluorescence image types (direct GFP: dGFP, GFP antibody: AbGFP, Arc antibody: AbArc) to compare GFP or Arc protein expression (red) and its respective control (grey): (a) Mean ROI-wide pixel intensity, (b) the standard deviation in ROI-wide pixel intensity (c) Mean soma-surround score, (d) the standard deviation in soma-surround score across cells, (g) the proportion of cells exceeding a 'high' soma-surround threshold. Data were averaged across all pixels or cells within an animal, then averaged across animals within a group. Results of a two-way t-test are indicated.

5.4 RESULTS

In this experiment the major aim is to examine the baseline expression of Arc in MEC layer II in preparation for assessing activity-dependent changes in the next chapter. To ensure that the level of Arc expression was at baseline, mice were housed in a quiet room for a week prior to the experiment and there were no disturbances to the room directly preceding removal of animals from their cage and their immediate subsequent sacrifice. In order to identify cells actively expressing Arc, I quantified GFP and Arc protein expression in heterozygous Arc^{+ve}/GFP^{+ve} mice, which express an enhanced GFP construct under the control of the endogenous Arc promoter. To achieve both a direct measure of GFP expression and a neuron-based measure with increased sensitivity, I first measured native GFP fluorescence in fixed, cryoprotected, but otherwise unprocessed, tissue and second, I co-stained adjacent sections with an anti-GFP antibody and a NeuN antibody.

To briefly outline the method, somatic expression was quantified by using a cell detection algorithm to detect cells both in direct GFP images and NeuN images, then calculating a soma-surround measure. For GFP/NeuN images this is the mean intensity of GFP fluorescence pixels within a region defined by a NeuN-detected cell with the mean intensity of GFP image pixels in the 'surround' of the NeuN-defined cell subtracted from it. High scores reflect higher levels of GFP within NeuN-detected cells relative to their immediate surround. For direct GFP and Arc protein images the cells first had to be detected using GFP/Arc protein expression before a score could be calculated. Cells are 'GFP^+' or 'Arc^+', first if detected by the cell detection algorithm, and second if the soma-surround exceeds a threshold score not detected in control tissue from a wild-type mouse.

5.4.1 *Baseline GFP expression follows a dorso-ventral gradient in MEC layer II*

RNA data from Allen Brain Atlas shows a gradient in Arc expression in MEC layer II so I first tested whether GFP follows a similar gradient using confocal images of GFP expression. I chose 7-8 regions of approximately 90 (width) x 180 (height) µm, spaced approximately 50 µm apart so that regions in which the layer was excessively thin could be excluded from the analysis, and quantified GFP expression across each region.

5.4.1.1 *Direct fluorescence measures*

Very few cells can be detected in MEC layer II using direct GFP fluorescence (Fig. 5.16). In 50% of the mice studied it is not possible to visually detect a single GFP^+ cell. The quantification of GFP fluorescence across animals confirms that very few cells can be detected (Fig. 5.19(B)). Given the size of each ROI, the maximum expected number of cells is just 2.4 per region. ROI-wide pixel intensity is low and shows no regional differences (Fig. 5.19(A)).

Analysis of the correspondence between distance from the dorsal MEC border and any ROI-wide pixel intensity or cell-based measure reveals no significant slope (Fig. 5.19(C)). Thus, due to the small number of detectable GFP^+ cells, I was unable to address the question of a possible gradient. However, this data indicates that baseline Arc expression in MEC is very low.

5.4.1.2 *GFP antibody measures*

To increase the detectability of the GFP signal, I used a GFP antibody, which increases the number of observable GFP^+ cells. However, cells showing high expression are sparse and unlikely to be present in more ventral regions (Fig. 5.18). ROI-wide pixel intensity and soma-surround measures also appear to decline across the group of mice (Fig. 5.18 (A-B)). The variability in soma-surround across cells is lower in ventral regions than dorsal regions (Fig. 5.18 (B)(b), 1-way ANOVA, Tukey HSD: $p<0.001$)

To investigate a systematic effect of dorso-ventral position, I calculated the linear regression slope for each measure, normalising all data points for each mouse. GFP antibody expression clearly shows a linear decline along the dorso-ventral gradient (Fig. 5.18 (C)(a), $slope = -1.4, p < 0.0001$). Interestingly, this gradient does not convincingly appear until $\sim 600\ \mu m$ from the dorsal MEC border. The limitation of ROI-wide measures is that they do not take into account changes in somatic area or density along the gradient. The more conservative measure of the mean soma-surround score of a cell is robust against light intensity differences across the coverslip and is independent of cell size or density. The mean soma-surround score and the proportion of all neurons that exceed a threshold soma-surround also decline along the gradient (Fig. 5.18 (C)(b-c), $slope = -0.74, p < 0.0001, slope = -0.15, p < 0.0001$). The pattern of changes for these measures in each mouse appears to be less linear than pixel intensity based measures, possibly due to the sparsity of high GFP-expressing neurons; just 3% of neurons in the most dorsal region and 0% in the most ventral region, on average, have a soma-surround score in the top 10% of the range (Fig. 5.18 (B)(d)). I performed a repeated-measures ANCOVA to test whether there is a robust gradient in individual mice. ROI-wide pixel intensity decreases along the gradient (Fig. 5.18 (C)(a): $F_{1,58} = 16.8, p = 0.0001$), but other measures do not show a significant reduction (Fig. 5.18 (C): $p > 0.05$).

5.4.1.3 *Neuronal statistics along the DV gradient*

To clarify the significance of this result, I performed the same analysis on neuronal statistics along the dorso-ventral axis of MEC layer II. NeuN pixel intensity and cell size do not vary along this axis (Fig.5.18(D), Intensity: $p = 0.59$, Area: $p = 0.72$), but neuron density decreases by approximately 1 neuron per 1000 μm^2 per 1mm in distance from the dorsal border (Fig.5.18(D), $slope = -0.97, p = 0.0017$). This reduction in density highlights the importance of using cell-based measures to establish a gradient rather than just ROI-wide intensity. However, using a repeated-measures ANCOVA, the reduction is not significant along the DV axis across individual mice (Fig. 5.18 (D)(b): $p > 0.05$).

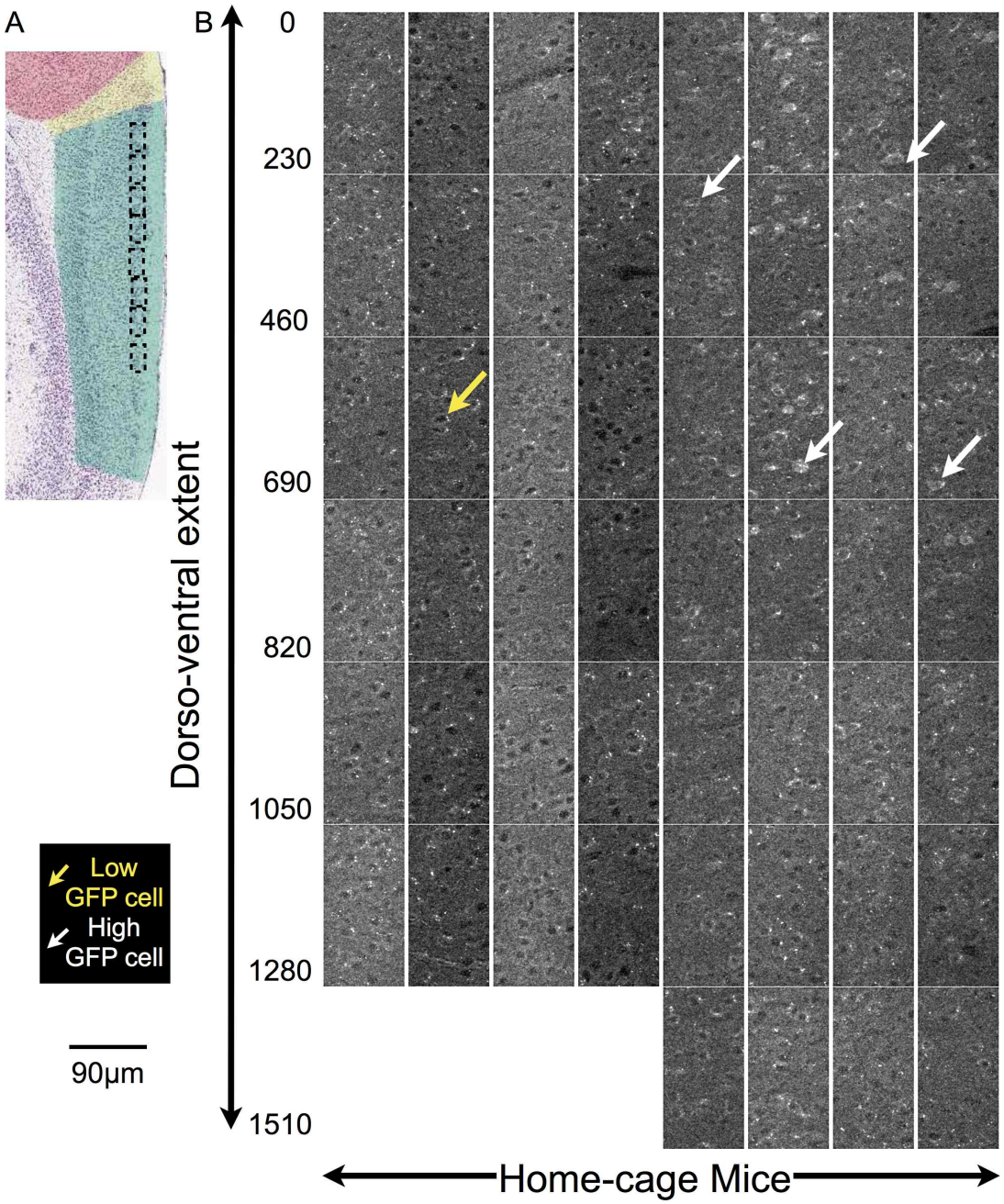


Figure 5.16: *Direct GFP fluorescence in MEC layer II.* (A) Nissl-stained sagittal section image of the MEC region (colour-coded green) to show the approximate location of the regions shown in (B). (B) Confocal 20x maximum projection images showing direct GFP fluorescence across all ROIs of MEC layer II from all home-cage animals (n=8). Scale (left) shows the approximate distance of the section from the dorsal border of MEC with parasubicular patches (in µm).

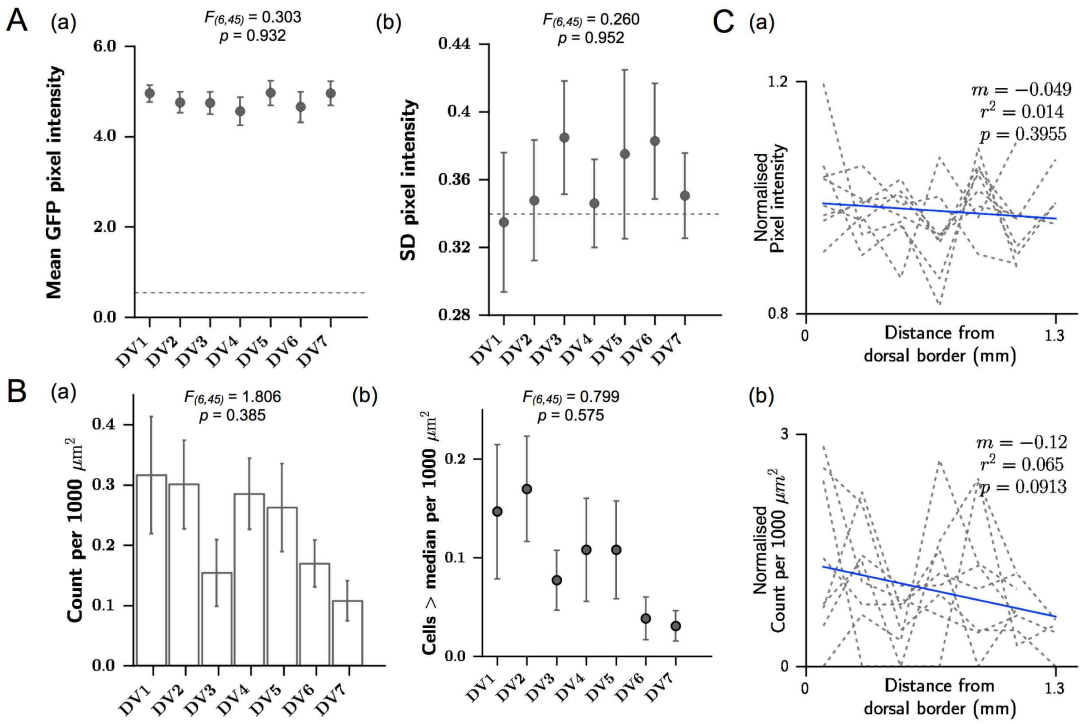


Figure 5.17: *Quantification of GFP expression along the DV gradient in mEC layer II.* (A) ROI-wide pixel measures for direct GFP: (a) Mean direct GFP fluorescence intensity, (b) Standard deviation in direct fluorescence intensity, (B) Cell-based measures for direct GFP: (a) Number of cells detected per 1000μm^2, (b) Number of cells detected that exceed the median soma-surround measure. The results of a 1-way ANOVA are shown. (C) Linear regression plots show (a) mean pixel intensity and (b) the number of cells directly detected, plotted as a function of distance from the dorso- ventral border. Blue line indicates the least squares regression fit. Grey dotted lines indicate data for each mouse. m denotes the slope for non-normalised values, r^2 the coefficient of determination, and p the significance of the gradient.

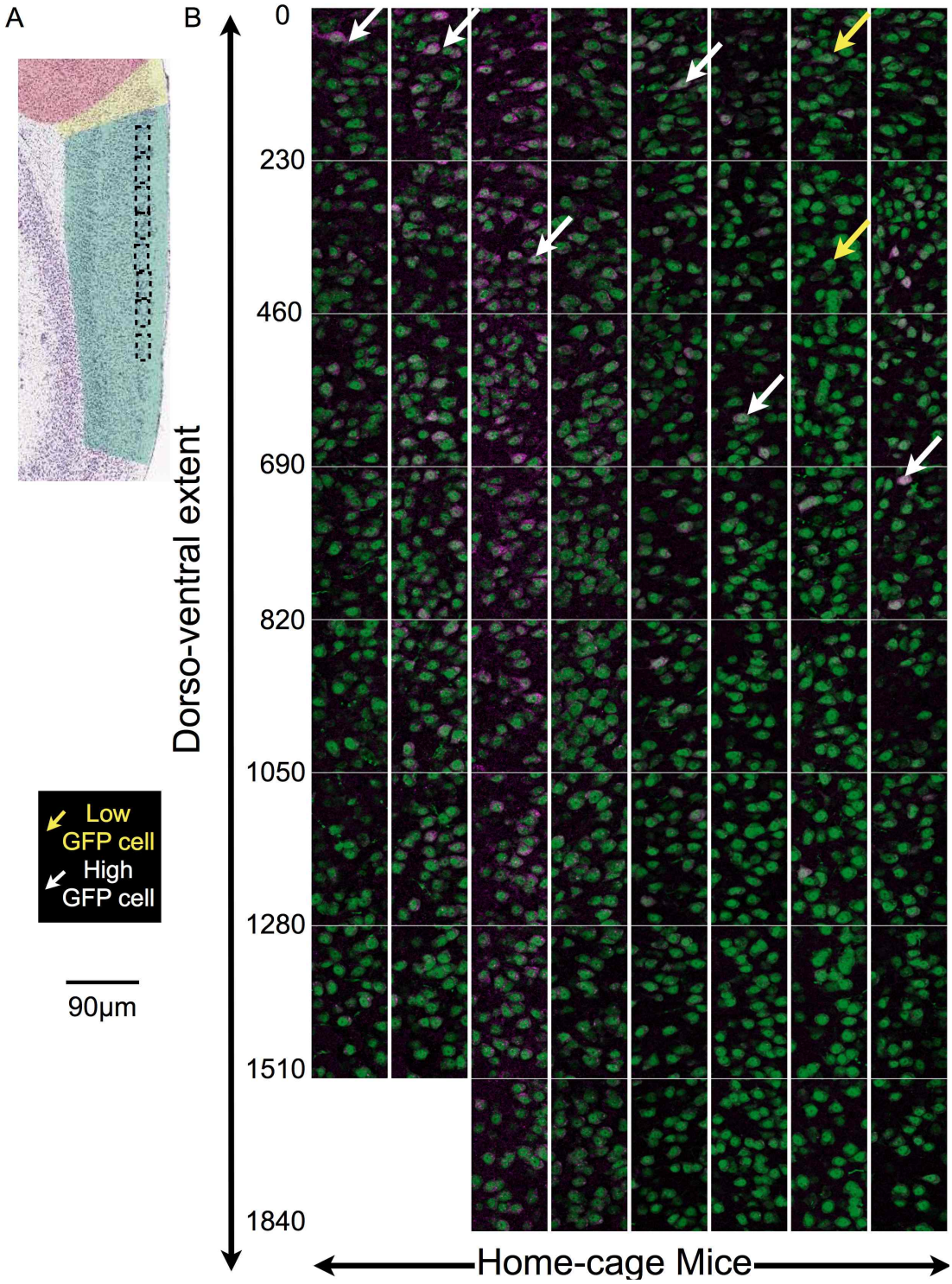


Figure 5.18: *GFP antibody expression in mEC layer II.* (A) Nissl-stained sagittal section image of the MEC region (colour-coded green) to show the approximate location of the regions shown in (B). (B) Confocal 20x maximum projection images showing the overlay of GFP antibody labelling (magenta) and NeuN labelling (green) across all ROIs of MEC layer II from all home-cage animals (n=8). Scale (left) shows the approximate distance of the section from the dorsal border of MEC with parasubicular patches (in µm).

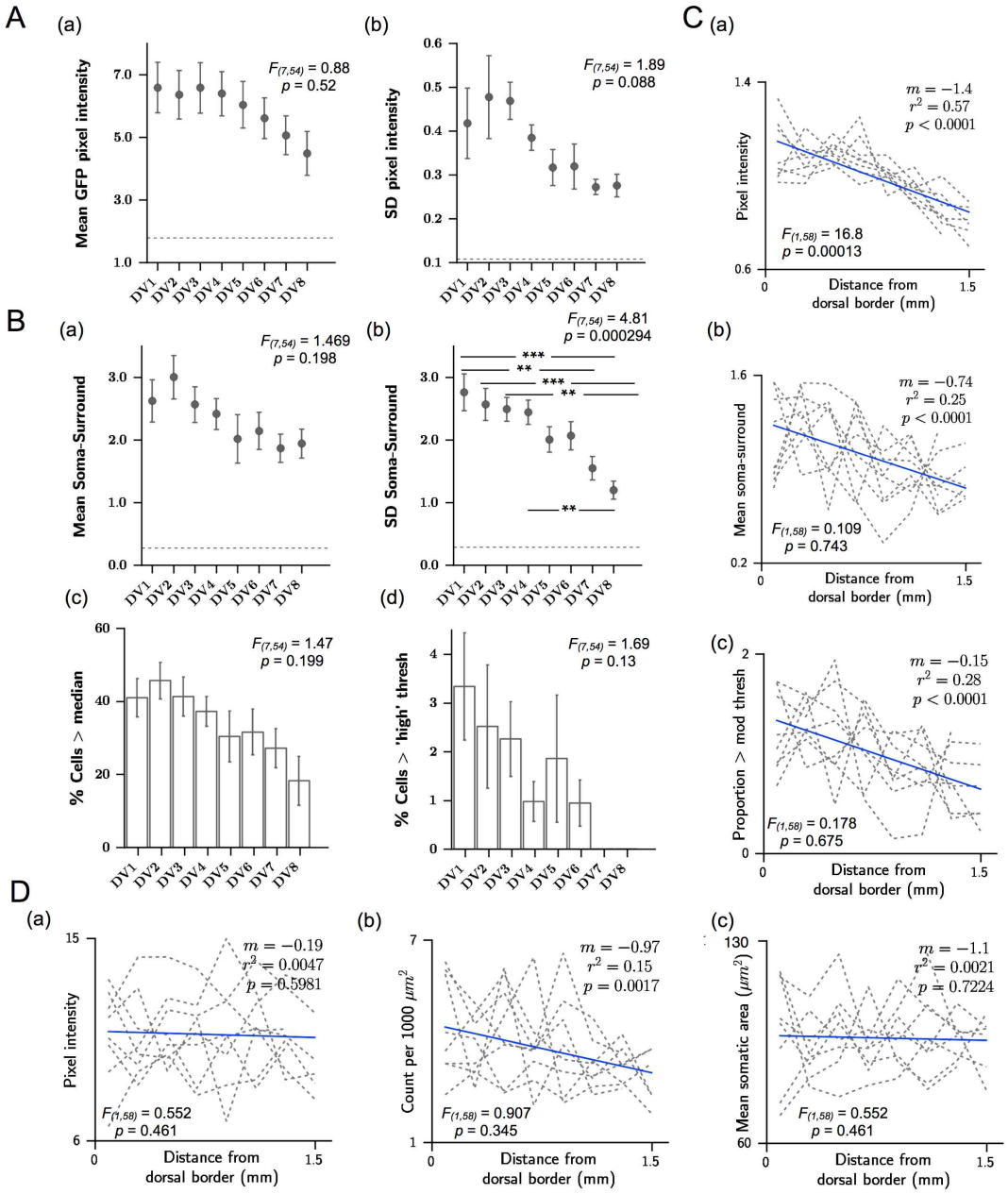


Figure 5.19: *Quantification of GFP antibody labelling reveals a DV gradient in mEC layer II.* (A) ROI-wide pixel measures: (a) Mean GFP antibody labelling pixel intensity, (b) Standard deviation in GFP antibody labelling pixel intensity. (B) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean soma-surround score, (b) standard deviation in soma-surround score, (c-d) Proportion of cells exceeding (c) median and (d) 'high' threshold soma-surround scores. The results of a 1-way ANOVA are shown, with significant post-hoc comparisons indicated for: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. (C) Linear regression plots showing normalised (a) ROI-wide mean pixel intensity, (b) mean soma-surround, (c) proportion of neurons exceeding the median, plotted as a function of distance from the dorsal border. (D) Linear regression plots show (a) Mean pixel intensity of NeuN labelling, (b) the number of NeuN-detected neurons per 1000 μm^2 and (c) the mean cross-sectional somatic area, plotted as a function of distance from the dorsal border. The significance of any slope was calculated using data normalised for each mouse. For linear regression plots grey dotted lines represent scores for each mouse. The blue line shows the line of best fit. m indicates the slope for non-normalised values. Results of a 1-way repeated-measures ANCOVA are shown in the bottom left of each axis.

5.4.2 *Baseline GFP expression in the MEC and other regions*

To put baseline GFP expression in MEC layer II in context, I compared it with expression in the deep layers of MEC, the large patches region immediately dorsal to the MEC, the hippocampal CA1 region and visual cortex, regions in which Arc expression has previously been studied or that exchange projections with MEC (regions indicated in Fig.5.20(A)).

Direct GFP fluorescence measures: Similar to MEC layer II, direct GFP fluorescence is very low in dorsal MEC deep layers (Fig.5.20(B)). A white arrow (i) indicates one of the few cells detected whereas yellow arrows indicate areas of the tissue that are devoid of fluorescence and mark the nuclei of non-GFP expressing cells. The GFP antibody labels a larger proportion of neurons, but as shown in the GFP/NeuN overlay, this is only a subset of the neurons, and few express high uniform levels of protein (Fig.5.20(B)). The group data show that the distribution of pixels is very similar across MEC regions (Fig. 5.21(A-B)). In addition, a similarly low number of cells can be detected in dorsal layer II and MEC deep layers and these cells appear to show similar properties, both to each other and other regions (Fig. 5.21(C-E)).

Direct GFP fluorescence also appears to be particularly low in CA1 (white arrows indicate low expressing GFP cells) (Fig.5.20(B)(iii)). Visual cortex deep and superficial layers are the only regions in which cells are clearly labelled with GFP that can be imaged directly, although the range of expression is high. Levels of direct fluorescence in the putative parasubicular large patches are also low.

Quantifying these direct GFP measures across all animals reveals regional differences in baseline GFP expression (Fig. 5.21 (A)). GFP is expressed at the highest levels in superficial visual layers, where direct fluorescence is significantly higher than in deep MEC layers and CA1 (Fig. 5.21 (B), $F_{5,42} = 4.21, p = 0.003, TukeyHSD : p < 0.01$). Nevertheless, ROI-wide pixel intensity data do not fully capture differences that can be observed from the confocal images, possibly due to the high noise level and very low fluorescence of GFP in regions such as MEC layer II. Cellular measures, however, are also limited for analysing direct GFP fluorescence due to the low numbers of detected cells across all regions (Fig. 5.21)(C-E)). I found no significant differences in the number of cells detected ($F_{5,42} = 1.18, p = 0.334$), the mean soma-surround of these cells ($F_{5,29} = 1.073, p = 0.396$), or the number of cells exceeding moderate or high thresholds ($F_{5,42} = 0.71, p = 0.618$,$F_{5,42} = 1.08, p = 0.384$).

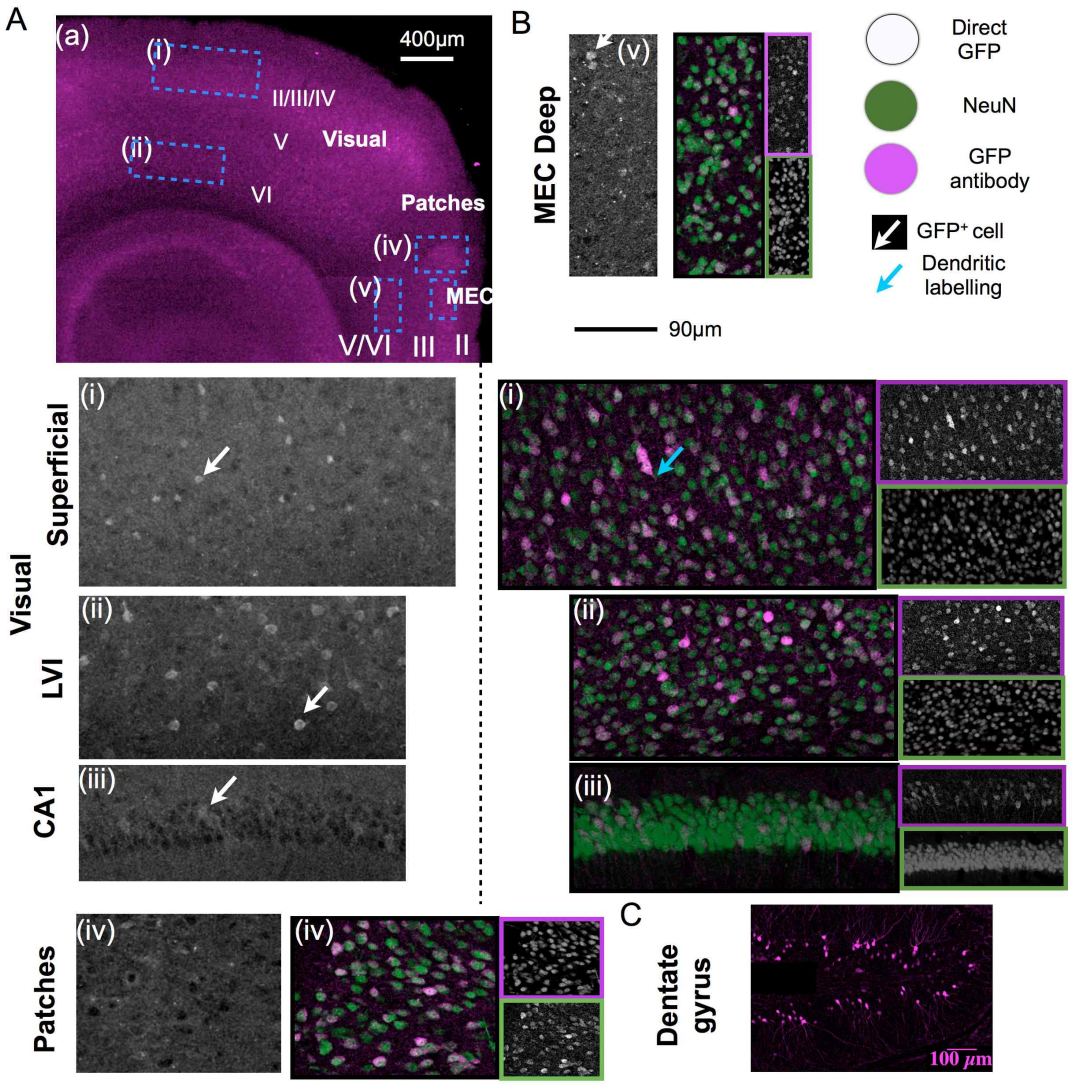


Figure 5.20: *Example images showing baseline GFP expression in different brain regions.* (A) (a) Stitched widefield image showing GFP antibody labelling at 4x magnification on a representative section. Blue boxes outline areas imaged at 20x magnification in (B): (i) Visual cortex superficial layers (II-IV), (ii) Visual cortex deep layer (layer VI), (iii) CA1 pyramidal layer (nb. actual image acquired more medially (Fig. 5.6)), (iv) parasubicular large patches, (v) MEC deep layers. (B) Grayscale confocal 20x images from a representative mouse show direct GFP fluorescence, and anti-GFP antibody (magenta outline), anti-NeuN labelling (green outline) and NeuN and GFP overlay (green/magenta) for each region. (C) Widefield image of GFP antibody labelling in the dentate gyrus, acquired at 4x magnification. DG labelling is reported in the next chapter.

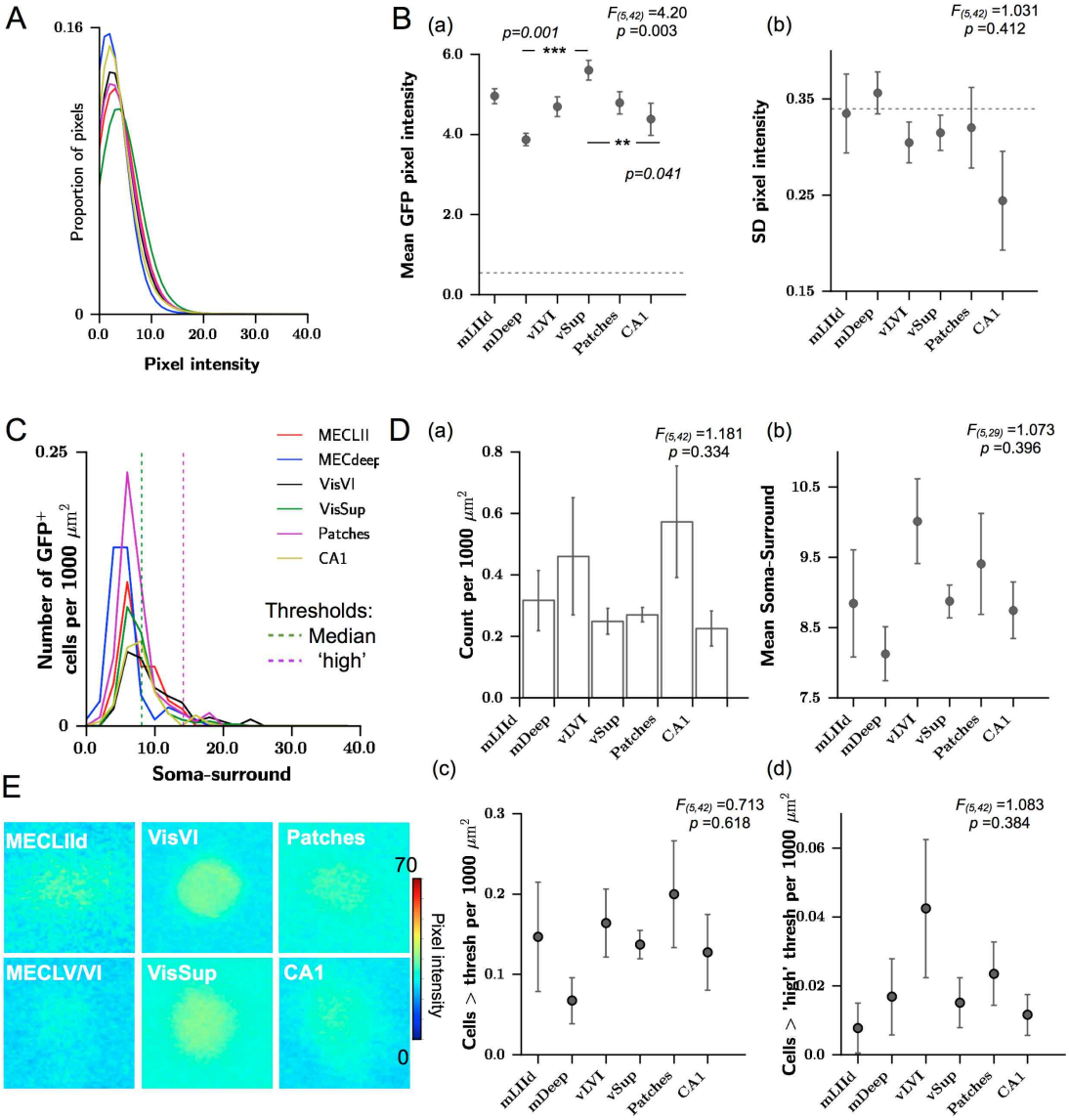


Figure 5.21: *Summary measures for baseline direct GFP fluorescence in MEC, visual cortex and CA1* (A) Histograms show the distribution of pixel intensities for directly imaged fluorescence in MEC deep layers, dorsal layer II, visual cortex (superficial and layer VI), the dorsal patches and CA1. (B) Directly-imaged GFP summary measures across the group of mice: (a) Mean pixel intensity across each region and (b) standard deviation in pixel intensity across each region.The grey dotted line shows the respective control result for the given measure. (C) Histograms show the distribution of soma-surround scores for GFP-expressing cells detected by direct fluorescence across regions. Dotted lines indicate the median and 'high' thresholds. (D) Summary measures for soma-surround score: (a) Density of cells detected per 1000 μm^2 (b) Mean soma-surround of cells across each region. nb. Data points are only plotted where > 3 GFP^+ cells were detected in the region and in many cases the number of cells used for analysis was very low (see (a)). (c-d) The density of cells per 1000 μm^2 exceeding the (c) median and (d) 'high' threshold. (E) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected neuron, calculated across all detected cells across all mice in each region. Results of a one-way ANOVA are shown (B,D), with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

GFP antibody measures: As was the case with the gradient in MEC layer II, use of the GFP antibody reveals differences across regions. Images of antibody labelling allow for detection of

strongly GFP-expressing cells, but these make up a very small proportion of cells (Fig.5.20(B)). Although there are no significant differences in ROI-wide pixel intensity across regions (Fig. 5.22(A-B), ($F_{5,35} = 1.81, p = 0.131$)), the soma-based measures clearly distinguish higher GFP expression in the visual cortical regions from MEC and CA1 (Fig. 5.22(C-E)).

Beginning with the MEC results, measures of ROI-wide pixel intensity (Fig. 5.22(A-B)), and the distribution of soma-surround scores for detected neurons is similar across dorsal layer II and V/VI (Fig. 5.22(C)). This can also be displayed visually by plotting an average neuron (Fig. 5.22(D)). The mean and variability in soma-surround score for neurons, and the proportion of cells exceeding low and high thresholds of GFP expression are also similar across MEC II and V/VI (Fig. 5.22(E),(Tukey HSD: $p > 0.05$)).

The features of MEC cells are typically indistinguishable from CA1 neurons, except for the proportion of CA1 neurons with supra-low threshold expression, which is significantly lower than all regions (Fig. 5.22(E)(c), ANOVA $F_{5,35} = 16.9, p < 0.00001, Tukey : p < 0.001$). The calculation for CA1 is necessarily different than from other regions because of the high cell density and difficulty in determining a 'surround', so this may contribute to this difference.

The features of MEC cells are also statistically indistinguishable from the patches dorsal to MEC, both using ROI-wide pixel intensity and somatic measures (Fig. 5.22, Tukey HSD: $p > 0.05$). However, there is some evidence that the cells in large patches express more GFP in their soma than MEC cells, but the variability between animals is quite high (Fig. 5.22(D-E), Tukey: $p = 0.123$). Patches neurons express more somatic GFP than CA1 neurons (Fig. 5.22(E)(a), Tukey HSD: $p < 0.001$).

Confocal images suggest that baseline expression in visual cortex is both higher and more variable across cells than in MEC. This difference is not significant at the level of ROI-wide pixel intensity (Fig. 5.22(A-B)), but the mean soma-surround score for superficial visual cortex neurons is significantly higher than for superficial MEC neurons (Fig. 5.22(C-E) (a), Tukey: $p < 0.01$). The difference in mean expression is closely matched by that for variability (Fig. 5.22 (E)(b) Tukey: $p < 0.001$), suggesting that in regions in which average cell expression is higher, this does not simply reflect a shift in the distribution of GFP expression, but a change in the way in which different types of cell up-regulate Arc. Variability in visual superficial cortex is also significantly higher than MEC deep layers (Fig. 5.22 (E)(b) Tukey: $p < 0.001$). The effect is that the proportion of high-expressing neurons is significantly higher in visual superficial cortex than either MEC region (Fig. 5.22 (E)(d) Tukey: $p < 0.01$). Neurons in layer VIa of visual cortex do not show statistically different GFP expression from MEC cells, but there is some evidence that somatic expression is more variable across cells (LV: $p = 0.065$, LII: $p = 0.098$), and that a larger proportion express high levels of GFP (LV: $p = 0.23$, LII: $p = 0.30$) (Fig. 5.22 (E)).

In summary, Arc-GFP activation appears to be relatively low in basal conditions in MEC, visual cortex, the large patches and CA1 and can only be quantified at the cellular level using a GFP antibody. The most consistent and high expression appears to be in visual superficial layers, although low-expressing cells are also observed there. Around ~80% of MEC cells show 'low' GFP expression, defined as more than would be expected from non-specific staining.

5.4.2.1 *Neuronal statistics across regions*

The use of NeuN as a neuronal marker enabled the quantification of the size and density of neurons across the MEC, as well as the other brain regions studied, thereby providing information that could provide insight into reasons for differences in the number of cells detected and ROI-wide pixel intensity across regions. The level of NeuN antibody labelling after subtraction of the 'background' intensity is highest, on average, in CA1 (Fig.5.23(a)). This is likely to be due to the unusual cytoarchitecture of the region and high cell density, and highlights the difficulty in comparing ROI-wide pixel intensity across regions. MEC layer II fluorescence is significantly higher than superficial visual cortex (Fig.5.23(a), Tukey: $p < 0.01$). Importantly, for comparisons of GFP^{+ve} cell density, neuron density is similar across regions ($F_{5,42} = 2.42, p = 0.051$, approx. 4 neurons per 1000 μm^2), with some evidence for densities up to 5-6 neurons per 1000 μm^2 in MEC layer V (Fig.5.23(b)). However, somatic size shows considerable differences ($F_{5,42} = 11.46, p < 0.0001$), with MEC layer II cells significantly larger than all other regions (Tukey HSD: $p < 0.01$, approx. 100 μm^2 mean cross-sectional area throughout the stack), and CA1 neurons significantly smaller (Fig.5.23(c)). MEC deep layer neurons are of similar size to neurons in other visual cortex and patch areas.

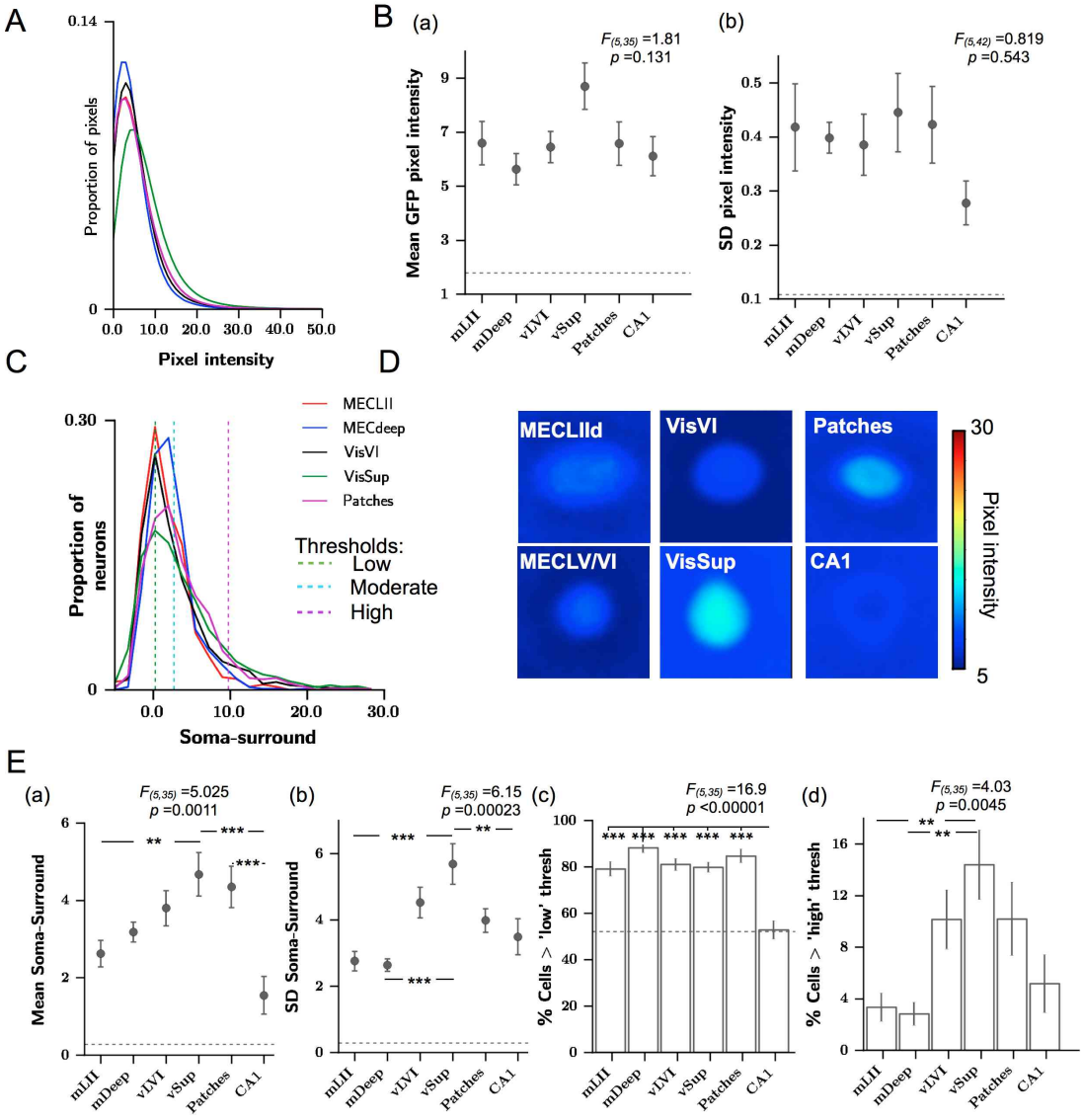


Figure 5.22: *Summary measures for baseline GFP antibody labelling in MEC, visual cortex and CA1* (A) Histograms show the distribution of ROI-wide pixel intensities for GFP antibody labelling in MEC deep layers, MEC dorsal layer II, visual cortex deep and superficial layers, the dorsal patches and CA1. (B) GFP antibody summary measures across the group of mice: (a) Mean pixel intensity, (b) standard deviation in pixel intensity.The grey dotted line shows the respective control result for the given measure. (C) Histograms show the distribution of soma-surround scores for all neurons detected using NeuN across regions. Dotted lines indicate the low, median and 'high' thresholds. (D) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron, calculated across all detected cells across all mice in each region. (E) Summary measures for soma-surround score: (a) Mean soma-surround of cells across each region. (b) standard deviation in soma-surround score, (c-d) The proportion of cells exceeding the (c) 'low' and (d) 'high' threshold. Results of a one-way ANOVA are shown (B,E), with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. n = 8

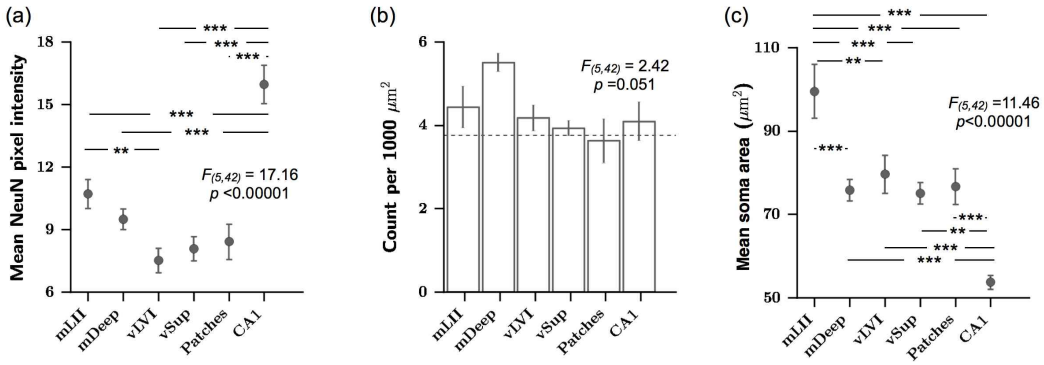


Figure 5.23: *Neuron density and size across regions.* A comparison of neuronal measures across MEC layer V/VI (mDeep), dorsal layer II (mLIId), visual cortex layer VI (vLVI) and superficial layers (vSup), the patches (Patches), and CA1. (a) Mean ROI-wide fluorescence intensity for NeuN immunofluorescence, (b) Mean detected neuron count per 1000 μm^2, (c) Mean cross-sectional cell area of detected cells (μm^2). Results of a one-way ANOVA are shown, with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$ (n=8).

5.4.3 *Arc protein expression in MEC and visual cortex*

To investigate expression of Arc protein, I also used immunofluorescence with a monoclonal antibody against Arc. To co-stain cells for viewing purposes I used NeuroTrace, a fluorescent Nissl stain. The aim was to address three questions: how does Arc protein expression level compare to GFP expression, how does protein expression compare in dorsal and ventral regions of MEC layer II, and how does protein expression compare across the MEC regions and visual cortex?

Initial inspection of confocal images suggests that Arc protein expression at baseline is difficult to detect in MEC layer II and the deep layers (Fig. 5.24). It is similarly low in visual cortex. I could not directly compare neuronal expression of Arc and GFP protein because NeuroTrace labels the majority of neural, glial and other cells, but the images suggest that cellular labelling of Arc protein is extremely sparse across all regions (Fig. 5.24).

I showed previously that protein labelling is likely to be present in visual cortex tissue as the distribution of ROI-wide fluorescence intensities is higher than in GFP^+/GFP^+ mice. Since Arc protein is targeted to synapses, it is possible that a large proportion of its baseline expression is in the neuropil rather than cell bodies. It is therefore important to consider both ROI-wide pixel intensity measures and somatic labelling in automatically and manually detected cells.

Since GFP protein expression follows a gradient in MEC, I directly tested whether dorsal Arc protein expression is higher than ventral Arc protein expression. Although the histograms for ROI-wide pixel intensity are overlapping, mean pixel intensity is higher in dorsal tissue (Fig. 5.25(A) and (B)(a), One-way paired t-tests $p = 0.03$, $n = 3$ pairs). However, there are no significant effects on other measures (Fig. 5.25, SD pixel: $p = 0.82$, Mean soma-surround $p = 0.44$, Cell count exceeding low $p = 0.17$ and high $p = 0.080$ thresholds). Changes in ROI-wide pixel intensity could simply be a result of the reduced number of cells in ventral regions. I did not explicitly question the presence of a gradient in Arc protein expression due to the small number of animals and difficulty in analysing expression within cells.

I had no *a priori* indication about relative levels of Arc protein in MEC layer II and V/VI and how this might compare to visual cortex deep layers, other than the subtle differences in GFP expression. Across MEC and visual cortex, there are no significant differences in mean or variability in ROI-wide pixel intensity between regions, despite an overall effect on the mean (Fig. 5.25(B), $F_{3,11} = 3.78, p = 0.04$) or in the number of cells exceeding a low or high threshold soma-surround score (Fig. 5.25(D)(b-c),$F_{3,11} = 0.73, p = 0.55$, High: $F_{3,11} = 0.861, p = 0.490$). However, dorsal MEC layer II detected cells appear to express higher levels of protein, relative to their surround, than MEC deep layers or visual cortex (Fig. 5.25(D)(a), $F_{(3,10)} = 7.64, p = 0.006, Tuk : p < 0.05$). In fact, on a number of measures MEC deep layer Arc protein expression appears to be indistinguishable from control Arc KO tissue (Fig. 5.25(B,D)).

To visualise this result, the Arc protein expression surrounding a detected cell has been plotted in Fig. 5.25(E). This result highlights that Arc protein expression forms a ring-like pattern around a detected cell, particularly in MEC layer II. It also shows that the average visual cortex cell and average dorsal cell measured, not taking into account differences across animals, actually express similar levels of Arc. Given that the number of animals used in this analysis was small (dorsal MEC layer II: 4, visual: 3), and the results quite variable, and that visual cortex and dorsal MEC do not differ on other measures, it is possible that this finding is not generalisable.

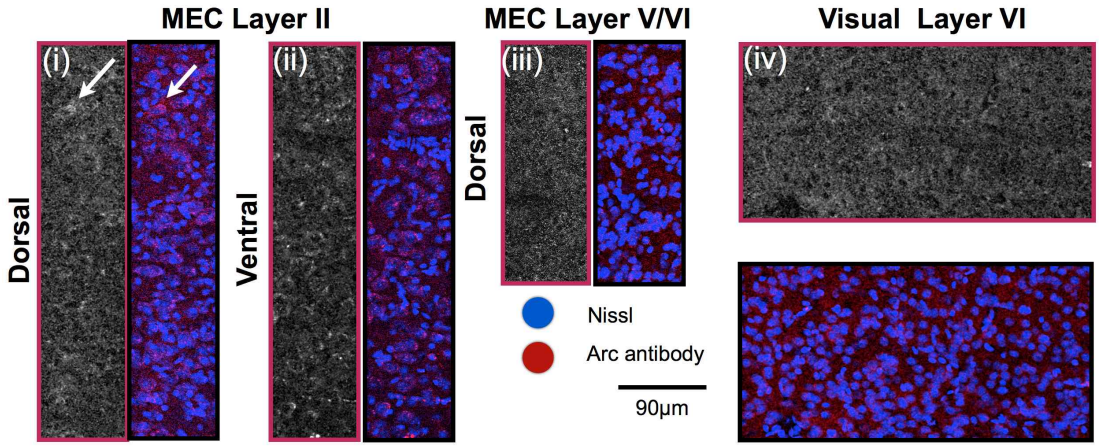


Figure 5.24: *Example images showing Arc protein expression in MEC and visual cortex.* Confocal 20x maximum projection images showing Arc antibody labelling (grayscale) and the Arc (red)/NeuroTrace(blue) overlay from a representative home-cage mouse tissue section. Regions shown are (i) dorsal layer II and (ii)ventral layer II, (iii) MEC dorsal deep layers, and (vi) visual cortex deep layer. White arrows show cells that express Arc.

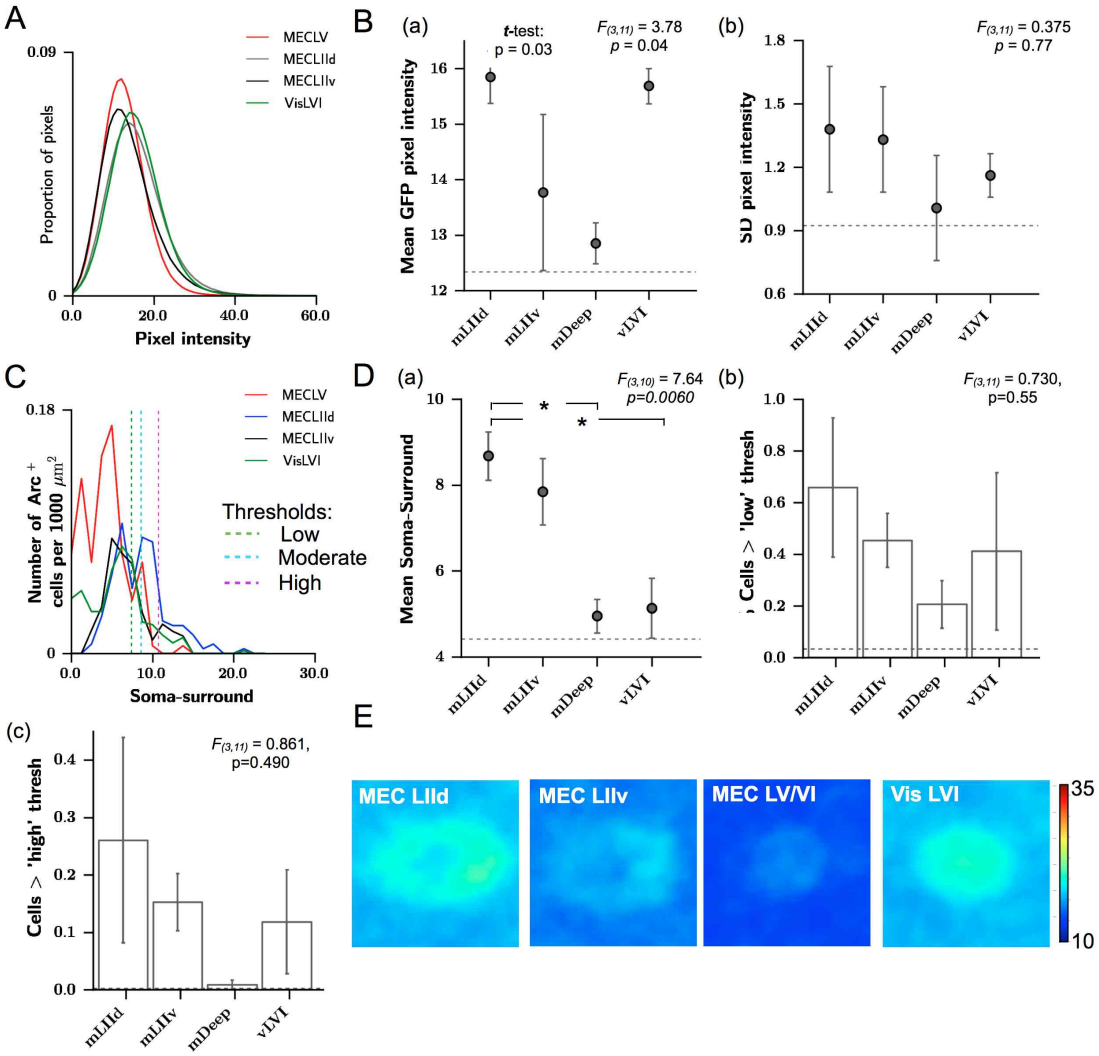


Figure 5.25: *A summary of Arc protein expression in MEC and visual cortex.* (A) Histograms showing the distribution of ROI-wide pixel intensity across MEC layer V, dorsal MEC layer II, ventral MEC layer II and visual cortex layer VI. (B) Summary measures for ROI-wide pixel intensity: (a) mean, (b) standard deviation. (C) Histograms showing the distribution of soma-surround scores for detected Arc-labelled cells. (D) Summary measures for Arc cells: (a) mean soma-surround, (b-c) number of cells exceeding a (b) low and (c) high threshold soma-surround score. (E) Images show the average 60x60 pixel region of Arc antibody labelling surrounding a detected cell, calculated across all detected cells across all mice in each region. Results of a 1-way ANOVA are shown, with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$ (MEC LIId: n = 4, LIIv:n = 3, LV: n=5, Vis: n = 3).

5.4.4 *GFP expression in GFP$^+$/GFP$^+$ Arc KO animals*

Arc protein performs a number of functions that may ultimately regulate Arc protein and mRNA levels (Bramham et al., 2010). To investigate whether removal of Arc protein affects the pattern of Arc activation in MEC at baseline levels, I measured GFP expression in GFP$^+$/GFP$^+$ Arc KO animals. Although this does not provide a direct comparison of expression because GFP levels are likely to be higher in GFP$^+$/GFP$^+$ than Arc$^+$/GFP$^+$ mice due to increased copy number of RNAs, it does enable a means of comparing relative expression within and across brain regions.

5.4.4.1 *Gradients in GFP expression in MEC layer II of GFP$^+$/GFP$^+$, Arc KO mice:*

GFP expression in MEC layer II of GFP$^+$/GFP$^+$ mice shows similar patterns to Arc$^+$/GFP$^+$ mice and is typically sparse in dorsal and ventral regions (Fig 5.26(A)). To address whether there is a gradient in expression in these animals, I used linear regression to examine the relationship between dorso-ventral distance and direct GFP fluorescence. ROI-wide pixel intensity decreases along the gradient ($slope = -1.4, p = 0.009$), as does the proportion of GFP$^+$ cells with threshold exceeding the MEC layer II median (Fig 5.26(B),$slope = -0.35, p = 0.0032$). However, these features may simply be the result of reduced cell density. Although a very weak gradient can be detected for ROI-wide pixel intensity using a GFP antibody ($slope = -0.85, p = 0.0127$), the mean soma-surround of cells does not change significantly along the gradient ($slope = 0.1, p = 0.6552$).

5.4.4.2 *GFP expression in Arc KO mice across MEC, visual cortex and hippocampus*

Direct GFP fluorescence measures Direct GFP expression in MEC deep layers appears to be similarly low to layer II and to expression in Arc$^+$/GFP$^+$ mice (Fig. 5.28). Note, that although data for both genotypes are shown together, they cannot be directly compared because of potential differences in imaging conditions. There are no ROI-wide differences in pixel intensity between the regions (Fig. 5.28(A-B)). However, quantification of cell-based measured reveals a significantly higher number of cells exceeding threshold expression in dorsal layer II than dorsal deep layers (Fig. 5.28(C-E)(c-d), Tukey, $p < 0.001$).

There are no ROI-wide differences in GFP expression across MEC regions, visual cortex and CA1 (Fig. 5.28(A-B), $F_{5,18} = 2.053, p = 0.119$). There are also no differences in the number of cells detected overall ($F_{5,18} = 2.053, p = 0.119$). However, the mean expression of cells in the large patches is significantly higher than the deep MEC layers, superficial visual cortex and CA1 (Fig. 5.28(E)(a), $F_{5,18} = 5.34, p = 0.0039, Tukey : p < 0.01$), and the proportion of cells exceeding a high threshold is significantly higher in MEC layer II than in CA1 or the deep visual cortex layers (Fig. 5.28(E)(d), $F_{5,18} = 9.32, p = 0.00016, Tukey : p < 0.01$)).

I made several observations about differences in genotype that could be important to directly test in a further experiment. Mean fluorescence appears to be higher in GFP$^+$/GFP$^+$ than in Arc$^+$/GFP$^+$ mice across all regions (Fig. 5.28(B)(a)). Variability in fluorescence intensity is also higher, but follows a different pattern, suggesting that differences in GFP expression across regions may not be linear (Fig. 5.28(B)(b)). I also compared soma-surround

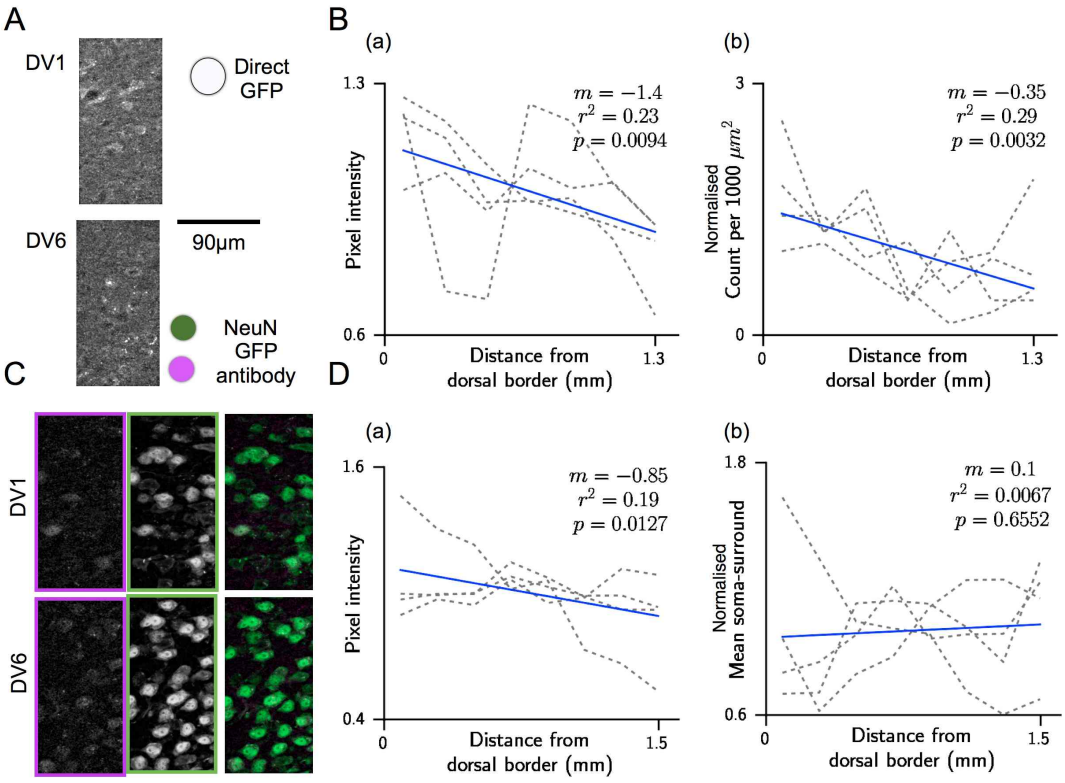


Figure 5.26: *Gradients in GFP expression in MEC layer II of GFP$^+$/GFP$^+$ (Arc KO) mice* (A) Example confocal 20x images showing direct GFP fluorescence in a dorsal and ventral MEC layer II region. (B) Linear regression plots for direct fluorescence of (a) normalised ROI-wide pixel intensity, and (b) normalised detected cell count per 1000 μm^2. (C) Example confocal 20x images showing GFP antibody labelling and NeuN labelling in a dorsal and ventral MEC layer II region. (B) Linear regression plots for GFP antibody labelling of (a) normalised ROI-wide pixel intensity, and (b) normalised mean soma-surround score. Dotted lines indicate results for individual mice. Blue line indicates the line of best fit, with slope m.

measures between GFP$^+$/GFP$^+$ and Arc$^+$/GFP$^+$ mice because this measure should be robust against laser intensity differences and, to some extent, staining efficacy (although this has not been tested statistically because there may be some effect). In superficial visual and MEC layers, an increased number of cells can be detected using direct GFP in GFP$^+$/GFP$^+$ than in Arc$^+$/GFP$^+$ mice (Fig. 5.28(B)(a)), consistent with an increased expression of GFP. Detected cells have higher soma-surround scores, but the increase is not uniform across regions, suggesting that it is not simply an effect of differences across experiment conditions.

GFP antibody measures Using an antibody against GFP and secondary detection reveals no significant differences across regions, using ROI-wide measures (Fig. 5.29(A-B)), or soma-based measures (Fig. 5.29(C-E)), except for a reduced proportion of CA1 neurons exceeding a low threshold soma-surround score (Fig. 5.29(E)(c), Tukey: $p < 0.01$). The pattern of results across regions is similar for Arc$^+$/GFP$^+$ and GFP$^+$/GFP$^+$ mice, but overall GFP expression appears to be consistently higher in GFP$^+$/GFP$^+$ mice. The reason for the difference in results between direct GFP fluorescence and GFP antibody labelling is not clear. One possibility is

that direct GFP fluorescence 'cells' are in fact regions of the neuropil where GFP protein is expressed in dendrites.

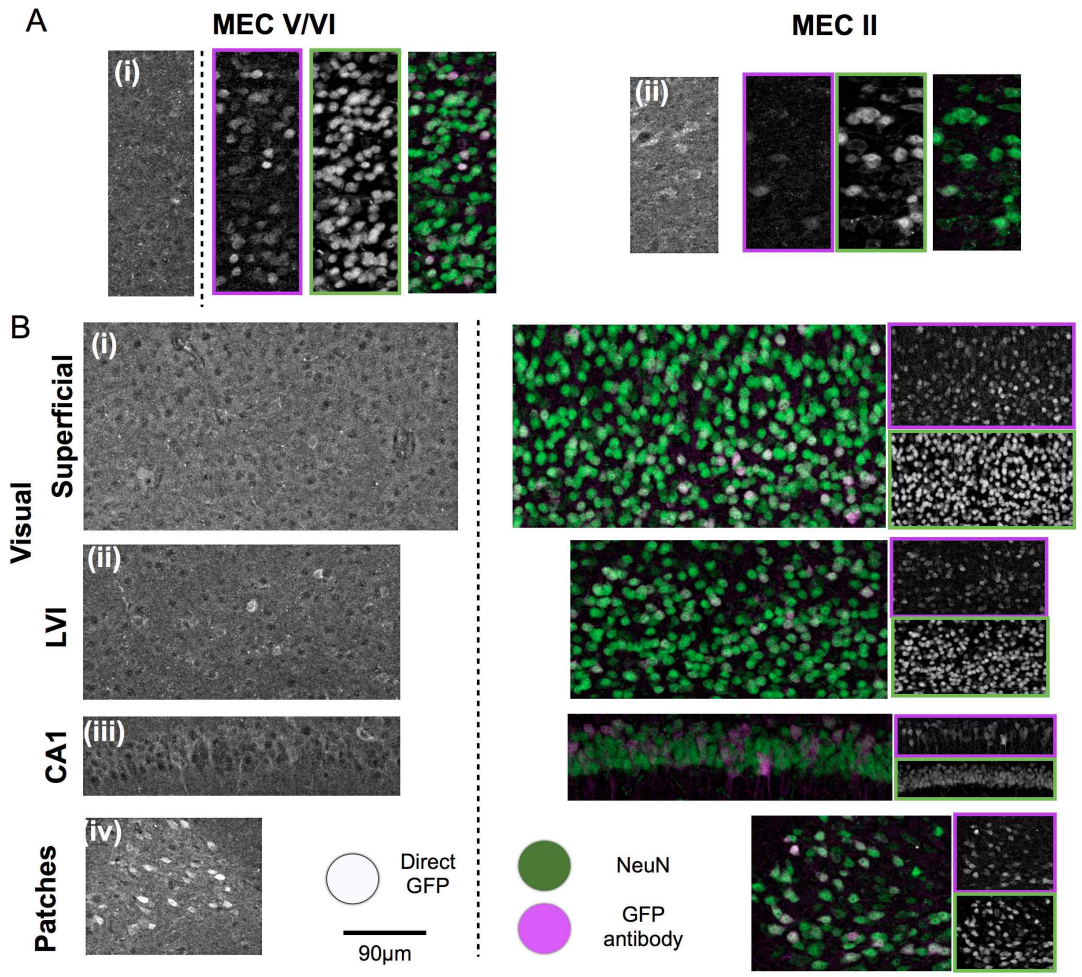


Figure 5.27: *Example images showing baseline GFP expression in a GFP$^+$/GFP$^+$ (Arc KO) mouse.* (A) Representative confocal 20x maximum projection images showing direct GFP expression (grayscale, no outline) from one section, and GFP antibody labelling (magenta outline), NeuN labelling (green outline) and GFP (magenta)/NeuN(green) overlay from an adjacent section. Data shown from (i) MEC deep layers, (ii) dorsal layer II. (B) As (A) for (i) superficial visual cortex, (ii) Visual cortex layer VIa, (iii) CA1, (iv) Patches.

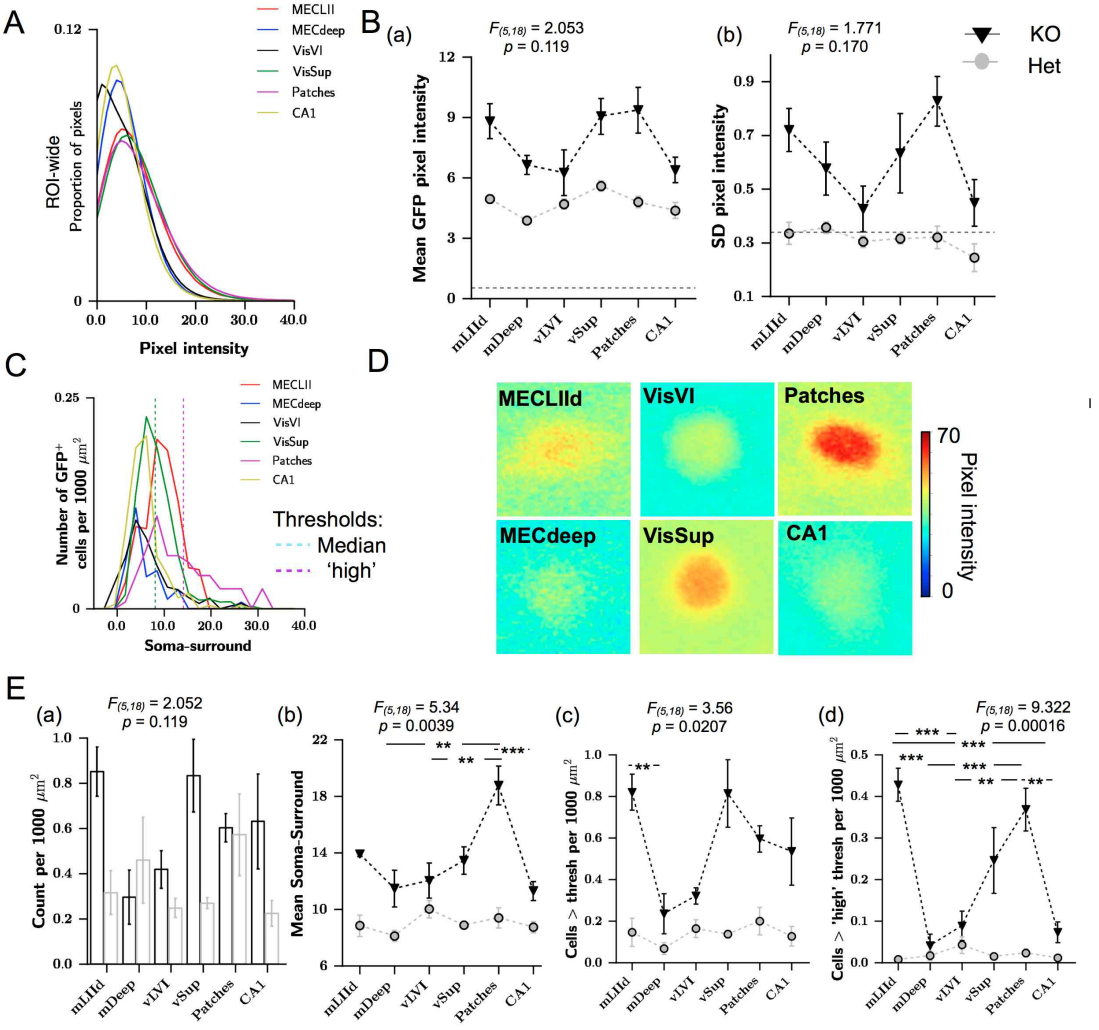


Figure 5.28: *Summary measures of baseline direct GFP fluorescence in GFP$^+$/GFP$^+$ mice.* (A) Histograms show the distribution of ROI-wide pixel intensities in MEC deep layers (LV: blue), dorsal layer II (LIId: red), visual cortex (LVI: black, superficial: green), the large patches (Patches: pink) and CA1 (yellow). (B) ROI-wide measures of: (a) Mean pixel intensity, (b) standard deviation in pixel intensity across each of the brain regions. (C) Histograms show the distributions of GFP$^+$ cells detected across different regions. (D) Images show the average 60x60 pixel region of GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice in each region. (E) Summary plots for GFP$^+$ cell-based measures: (a) the density of cells detected, (b) mean soma-surround for detected cells, (c-d) the density of cells per 1000 µm^2 exceeding the (c) median and (d) 'high' threshold. Results of a one-way ANOVA are shown, with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. Equivalent results are shown for Arc$^+$/GFP$^+$ animals (light grey) for comparison purposes. These differences were not statistically tested because experiments for the two groups were completed in different sessions.

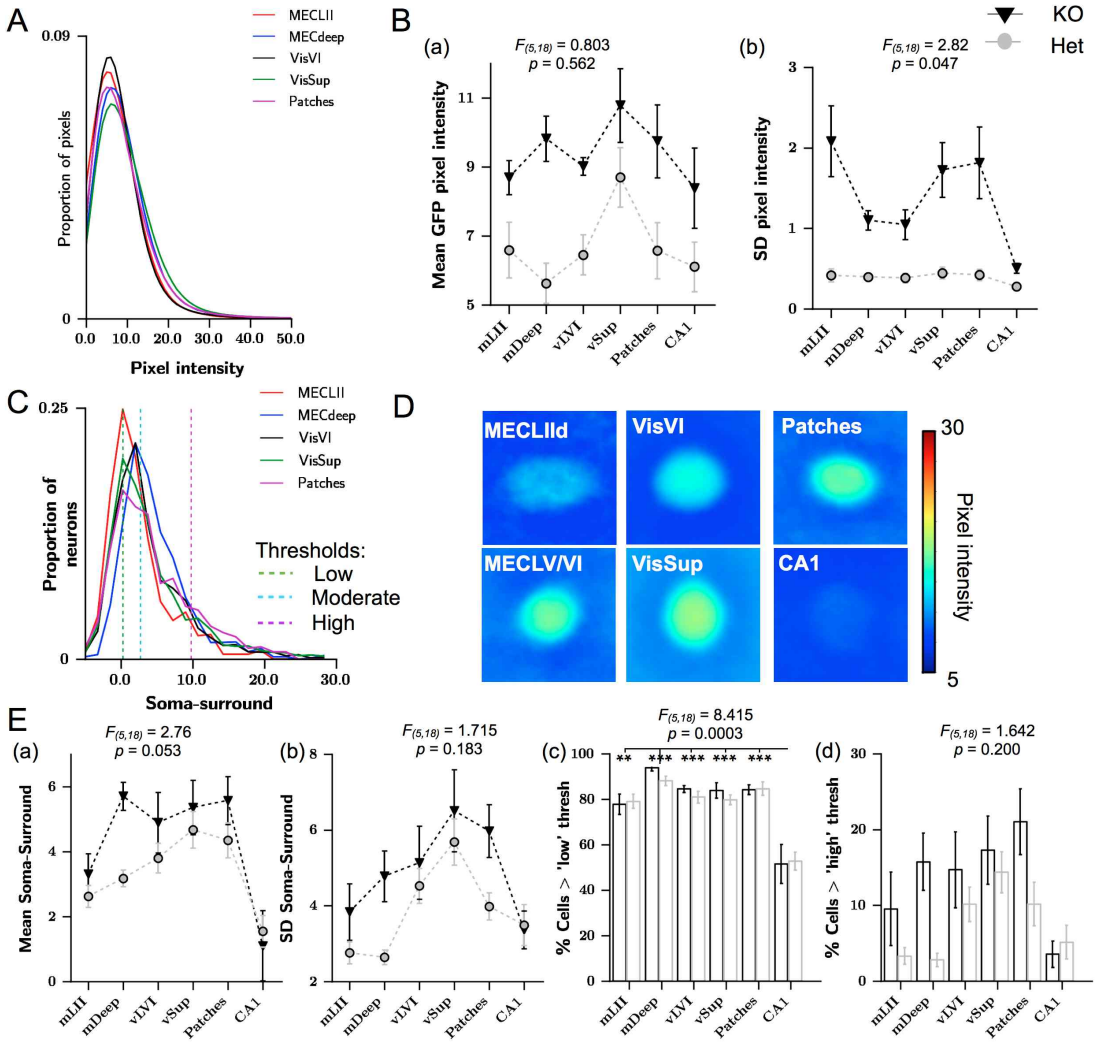


Figure 5.29: *Summary measures of baseline GFP antibody labelling in GFP$^+$/GFP$^+$ mice.* (A) Histograms show the distribution of ROI-wide pixel intensities in MEC deep layers (LV: blue), dorsal layer II (LIId: red), visual cortex (LVI: black, superficial: green), and the large patches (Patches: pink). (B) ROI-wide measures of: (a) Mean pixel intensity, (b) standard deviation in pixel intensity across each of the brain regions. (C) Histograms show the distributions of soma-surround scores for all neurons. (D) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron, calculated across all detected cells across all mice in each region. (E) Summary plots for neuron-based somatic measures: (a) mean soma-surround for all detected neurons, (b) the standard deviation in soma-surround score, (c-d) the proportion of neurons exceeding the (c) 'low' and (d) 'high' threshold. Results of a one-way ANOVA are shown, with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. Equivalent results are shown for Arc$^+$/GFP$^+$ animals (light grey) for comparison purposes. These differences were not statistically tested because experiments for the two groups were completed in different sessions.

5.5 DISCUSSION

5.5.1 *Summary*

For this experiment I developed a set of tools for the somatic analysis of Arc expression in heterozygous Arc^{+}/GFP^{+} and GFP^{+}/GFP^{+} mice. I used these tools to explore the spatial localisation of Arc expression in the MEC, hippocampus, visual cortex and parasubicular large patches. I show that basal levels of Arc-GFP activation (Wang et al., 2006) and Arc protein expression (measured using an antibody against Arc) are very low in MEC, particularly at the level of somatic expression. Nevertheless, there is a weak but detectable gradient in Arc-GFP activation within neuronal somata throughout the dorso-ventral axis of MEC layer II. There are no clear-cut differences in expression between MEC and other brain regions but subtle differences emerge when investigating somatic neuronal expression and the prevalence and expression levels of low- and high-expressing Arc cells. The superficial layers of visual cortex express significantly higher levels of Arc at baseline than MEC cells.

5.5.2 *The dorso-ventral gradient in the expression of Arc*

In situ hybridisation images from in the Allen Brain Atlas project reveal a gradient in the expression of Arc RNA in MEC that is particularly pronounced in layer II (Lein et al., 2007). This gradient is consistently found across animals sacrificed under task-free conditions and after sleep deprivation, and is persistent across light and dark cycle sacrifice (Thompson et al., 2010).

In this experiment, I show using linear regression analysis that there is a linear relationship between distance from the dorsal border and the mean level of Arc-GFP activation. Dorsal regions show higher expression, on average and within neurons, and an increased proportion of neurons express high levels of Arc. For most cell-based measures, this pattern only emerges when using a linear regression analysis, which assumes that errors on each data point are uncorrelated and which does not take into account that multiple data points are collected from each mouse leading to pseudoreplication errors. The gradient in Arc-GFP activation within cells is not consistent within individual sections for individual mice and thus does not reach significance using a repeated-measures ANCOVA. These observations indicate that if this gradient exists, which is consistent with the RNA gradient in Allen Brain Atlas data, then it may only have an effect at the level of a much larger population of cells than can be viewed in a single section. Such a gradient parallels the changes in the physiological properties of layer II cells (Garden et al., 2008), but it is not clear that it maps on to changes in grid cell size and spacing (Hafting et al., 2008; Fyhn et al., 2008), which are clearly evident within 1mm from the dorsal border. In addition, grid spacing follows a modular organisation (Stensola et al., 2012), such that an increased number of grid field modules exist in ventral regions than dorsal regions. This leads to the prediction that ventral cells should be more variable than dorsal cells, but I show that the variability in Arc-GFP activation between cells decreases along the dorso-ventral axis.

If Arc-GFP activation is indeed higher in dorsal regions then what is the functional significance ? Arc expression is considered to be dependent on activity (Lyford et al., 1995) so one possibility is that increased Arc expression reflects higher basal activity in dorsal cells. Differences in intrinsic excitability have not been reported in *in vitro* work (Garden et al., 2008), but this has not been explicitly addressed using whole-cell recordings in awake, behaving animals. One difficulty with testing this is that there is high variability in Arc expression between cells in the same region, so a large number would need to be sampled to establish such an effect. Alternatively, increased excitability could reflect the fact that the mice are housed in a small environment, leading to an increased probability of exciting the small but more numerous receptive fields of dorsal cells than the larger receptive fields of ventral cells.

Arc expression is also considered to be a marker of plastic changes, since Arc modifies synaptic properties (Chowdhury et al., 2006; Tzingounis and Nicoll, 2006) and is necessary for certain forms of long-term potentiation and depression (Plath et al., 2006). Increased expression in dorsal regions may therefore reflect increased capacity for plasticity. As the extent to which synaptic properties and plasticity change along the gradient has not been explicitly tested, it is difficult to speculate about how Arc might modulate this. Furthermore, since the animals studied here had no task to perform, it is not clear what the requirement or effect of increased plasticity might be in dorsal cells. I did not find a clear effect of region on Arc protein. Since it is the protein that influences the structural changes that plasticity may rely on (Bramham et al., 2010), it will be necessary to clarify this to establish what any differential effect on dorsal and ventral neurons might be.

5.5.3 *Differences in Arc expression between the superficial and deep layers of MEC*

MEC layer II and layer V/VI project to and receive inputs from different regions and contain different cell types (summarised in (Canto et al., 2008)). Perhaps most significantly, layer II provides the major input to the hippocampus, via the dentate gyrus and CA3 (Dolorfo and Amaral, 1998b; Schwartz and Coleman, 1981; van Groen et al., 2003), whereas layer V/VI receives a projection from the hippocampus, via CA1 pyramidal cells (Swanson and Cowan, 1977; Burwell and Amaral, 1998). Arc-GFP activation is similarly low in the dorsal region of layers II and V/VI. However, Arc protein levels in layer V are similar to Arc KO tissue containing no Arc protein at all. This suggests that although Arc is actively transcribed in a small proportion of layer V/VI cells under standard environmental conditions, Arc protein expression has different requirements to GFP expression. Dorsal layer II cells express higher levels of Arc protein than layer V/VI, which suggests that at least a small proportion of cells actively translate Arc under baseline conditions. Not only is this protein expressed in the soma, but also appears to be active in the neuronal projections. It is possible that movement around the home cage produces some requirement for Arc in the spatially modulated cells of layer II whilst layer V cells remain relatively inactive. This is supported by work using juxtacellular recordings in freely moving naive mice, which showed that layer V cells remain relatively silent whilst layer II stellate cells modulate their pattern of firing depending on location (Burgalossi et al., 2011). It is possible that without training or a requirement for spatial memory, CA1, and therefore layer V/VI, show very little Arc-dependent activity.

The expression of Arc has previously been examined under a number of conditions with results very similar to those presented here. The small amount of work reporting data in MEC showed that under baseline conditions, Arc RNA levels in rats are low (Gusev et al., 2005). There are no available images of Arc protein expression in MEC.

5.5.4 *Differences in Arc expression between the MEC and other brain regions*

Superficial visual cortex cells express Arc at higher levels than MEC regions. The researchers who developed the Arc-GFP line also examined baseline Arc-GFP activation in visual cortex (Wang et al., 2006). Although they showed almost no expression of Arc across the layers of visual cortex, their 'baseline' animals were sacrificed during the dark cycle. This suggests that residual GFP expression in the superficial layers in this particular study results from the presence of light and that Arc expression may be activated in a subset of cells just through the need to process visual information.

In the present study, I showed that CA1 pyramidal cells show low levels of Arc-GFP activation at baseline. This is consistent with work by Messaoudi et al. (2007), who showed no detectable Arc RNA or protein in the hippocampus of home-cage rats. A similar result was obtained by (Kubik et al., 2012) for caged rats. It should be noted that in images of Arc RNA expression from the Allen Brain Atlas, Arc RNA appears to be expressed at high levels in CA1 (Lein et al., 2007). These differences may result from the use of different probes, differences between the species or breed of animal, or may suggest that animals used to obtain ABA data were not sacrificed under truly baseline conditions, which is possible since methods did not explicitly aim to prevent immediate early gene expression or ensure that animals were not exposed to novel or stressful conditions immediately before sacrifice. Arc RNA is transcribed within minutes, peaking at 30 minutes in the cytoplasm (Lyford et al., 1995), so even being moved to a new room shortly but not immediately before sacrifice could lead to a sufficient increase.

The delineation of the MEC at the dorsal extent is not clear-cut and it has recently been suggested that the large patches of cells, evident with cytochrome oxidase staining (Burgalossi et al., 2011) may not be part of the MEC, but part of the parasubiculum. To my knowledge, Arc expression has not previously been quantified in these parasubicular patches. Does it support the view, as Burgalossi et al. (2011); Boccara et al. (2010) have posited, that the large patches contain a distinct neuronal subpopulation to the rest of the MEC? I found that large-patch neurons are significantly smaller than dorsal MEC layer II neurons, with an average cross-sectional area of 85 μm^2 , compared with $100 \mu m^2$ in dorsal MEC (note that this is an average cross-sectional area across all detected two-dimensional objects rather than the cross-sectional area through the largest section). However, I found no significant differences between Arc-GFP activation in the large patches and in MEC at baseline, although large patch expression appeared consistently higher.

5.5.5 *Limitations*

There are several potentially significant limitations of this experiment: (1) the use of heterozygous Arc-GFP animals that do not possess wild-type levels of Arc protein, (2) the use of GFP protein expression as a measure of Arc promoter activity in Arc-GFP mice, (3) the inability to analyse the types of cells expressing Arc, and (4) the lack of control for fast protein translation in dendrites.

The use of heterozygous Arc-GFP animals has two possible problems relating to its reduced levels of Arc protein. First, the phenotype of the animals is not that of a wild-type (WT) animal. There is some evidence that Arc^{+}/GFP^{+} mice may have increased susceptibility to seizures, as cortical extracellular recordings have revealed generalized patterns of sharp, synchronous epileptiform discharges that are similar to those of GFP^{+}/GFP^{+} mice (Peebles et al., 2010). Although spontaneous cortical seizures have not been observed under experimental conditions (Peebles et al., 2010), this could indicate that additional mechanisms not present in wild-type animals are required to control the excess excitability. Further to this, it is possible that the memory systems of animals with reduced levels of Arc could be altered due to its importance in memory function. This would be reflected in impairments on entorhinal or hippocampus-dependent tasks. Although such deficits have not been described in either heterozygous or homozygous Arc-GFP mice, this may be because they have not been adequately tested for, so I cannot be certain of their absence. Probing this system experimentally is therefore not necessarily equivalent to probing a wild-type system.

The second related issue is that since there is evidence that Arc may indirectly regulate subsequent expression of Arc and other proteins (Rao et al., 2006; Bramham et al., 2010), the level of Arc expression may not be representative of level in wild-type animals. If Arc regulation is similar across all cell types in the cortex and hippocampus then the comparison across regions presented here is valid, but we cannot necessarily make this assumption. However, I found a similar pattern of results for GFP expression in Arc^{+}/GFP^{+} and GFP^{+}/GFP^{+} mice, which suggests the loss of Arc protein does not dramatically change the system. Performing and comparing steady-state ISH and immunohistochemistry analyses in wild-type and Arc-GFP mice would provide an indication of whether the Arc promoter is indeed active and translation is occurring at wild-type levels at baseline.

Another potential limitation is that some of the conclusions of this experiment rely on the assumption that GFP expression is a valid proxy for reporting Arc promoter activity. As highlighted in Fig. 5.3, GFP protein is translated from RNA through a number of translational, and potentially post-translational mechanisms. If GFP translation can be suppressed, as Arc translation can be suppressed (reviewed in (Bramham et al., 2010)), then it may not be possible to detect all cells in which Arc is activated. The translation of GFP could also potentially be regulated in the absence of promoter activity via mechanisms similar to Arc. In addition, these processes may not be equivalent across cell types, which could lead to different levels of GFP protein with different peaks in expression, given the same level of RNA. A final point is that the GFP transgene in Arc-GFP mice contains a neo cassette with strong promoter activity that could influence transcription of the transgene, thereby reducing the ability of GFP protein to report on activity. It could also affect the transcription of neighbouring genes. Unexpected

effects on phenotype have previously been linked to the presence of neo cassettes (Olson et al., 1996). Removal of the neo cassette would alleviate this potential problem. However, previous research using these mice has not reported a problems relating to this (Wang et al., 2006).

Taken together, the potential limitations of this genetically modified mouse line suggest that separate measurements of RNA and protein expression in wild-type mice might provide a more accurate indication of Arc expression. However, this method would not provide the potential to simultaneously measure the distribution of cells in which the Arc promoter is active, independent of whether it is successfully trafficked and locally translated, and in which Arc protein is being translated. If GFP protein does provide a dynamic indicator of Arc promoter activity, as reported by (Wang et al., 2006), then use of Arc-GFP mice provides this potential, but further work will be necessary to ascertain how precisely GFP is trafficked, translated and regulated in MEC cells.

Although this analysis technique provides an indication of cellular somatic expression of GFP and Arc, the information is limited, particularly for Arc where NeuroTrace (a fluorescent Nissl marker) was used as an indicator of cells. Some cells stained by NeuroTrace are non-neuronal, including glial and endothelial cells (Quinn et al., 1995) because the marker stains the rough endoplasmic reticulum, which is present in all cells. Even when using NeuN, it is not possible to identify the type of cell or whether it is excitatory or inhibitory, in order to speculate as to the role played by Arc. Co-labelling tissue with interneuron markers or cell-specific markers could provide insight into this.

Finally, although the somatic protein expression results presented are likely to reflect baseline expression, dendritic Arc protein can be translated very quickly in certain cell types, inducing LTD (Waung et al., 2008). Although animals were typically perfused within 10 minutes of the initial anaesthesia, this could be sufficient to induce Arc protein expression in the neuropil.

5.5.6 *Conclusions*

Consistent with my predictions, Arc expression is low in the MEC at baseline. In the next chapter I explore how exposure to a novel environment influences the level of Arc expression in MEC. In particular, I'll consider whether Arc-related information processing differs across grid cells in layer II that encode different spatial resolutions by exploring whether novel exploration modulates the gradient. I'll also consider how Arc-related changes differ across the MEC, and whether the MEC, hippocampus and visual cortex follow similar patterns of changes.

6

THE EFFECT OF NOVEL EXPLORATION ON ARC EXPRESSION IN MEC, VISUAL CORTEX AND HIPPOCAMPUS

6.1 SUMMARY

In the preceding chapter, I investigated the baseline expression of the activity-regulated cytoskeleton-associated gene, *Arc*, across the medial entorhinal cortex (MEC) and other regions involved in spatial navigation. This gene is not only activity-dependent (Lyford et al., 1995), but is critical to certain forms of plasticity (Plath et al., 2006). In order to provide insight into how information relevant to the modulation of *Arc* levels is processed in spatial navigation circuits during a new experience, I placed animals into a novel environment containing objects and allowed them to explore. In this chapter, I outline where and how novel exploration influences *Arc* expression and summarise how this might be useful for understanding the MEC circuit.

1. Novel exploration does not cause up-regulation of *Arc* -GFP activation in MEC layer II despite the fact that these cells are known to be active during spatial exploration. It also has no effect on *Arc* -GFP activation in CA1 or the patches dorsal to MEC and a very modest effect, if any, in the dentate gyrus.
2. Behavioural up-regulation of *Arc* is more pronounced in the dorsal region of layer V/VI of MEC, a region that projects to many areas of neocortex, than in any other region of medial entorhinal cortex, visual cortex or hippocampus.
3. In the visual cortex, *Arc* -GFP activation modestly increases following exposure to a novel environment, which is consistent with previous work. However, the difference between home-cage animals and those allowed to explore is less robust than in layer V/VI of MEC.
4. Changes in *Arc* -GFP activation and Arc protein expression show broadly similar patterns; where *Arc* -GFP activation increases following exposure to a novel environment, Arc protein expression follows. However, in MEC layer II Arc protein level increases despite no detectable increase in Arc-GFP activation.
5. *Arc* protein is not necessary for the increase in *Arc* -GFP activation following exposure to a novel environment. In the majority of regions, including MEC deep layers, visual cortex and CA1, GFP^+/GFP^+ mice, which possess no *Arc* protein, show the same patterns of changes as Arc^+/GFP^+ mice.

These results lead to the conclusion that neurons in layer V/VI of MEC show a particular requirement for increased levels of Arc when an animal is placed into a new environment in which learning might occur. They lead to the prediction, to be tested in the next chapter, that the major hippocampal input from CA1 to MEC deep layers is not necessary for such plasticity-related changes.

6.2 INTRODUCTION

Neural circuits in the medial entorhinal cortex (MEC) can represent an animal's location in space as it explores a novel or familiar environment (Hafting et al., 2005; Moser et al., 2008). When the environment changes, this representation can dynamically remap whilst preserving important features (Barry et al., 2007; Fyhn et al., 2007). This re-mapping also occurs in hippocampal neurons (Fyhn et al., 2007), which suggests that important plasticity mechanisms may be occurring in the hippocampal-entorhinal system. The precise identity of the neurons involved and the mechanisms that underlie their behaviour remain unclear, but it is likely that there is an interaction between the integrative properties of cells and their anatomical organisation within MEC layers and hippocampal subregions.

As discussed previously in this thesis, gene expression provides an insight into the integrative properties of neurons. I have shown in the previous chapter that the activity-regulated gene Arc is expressed in the MEC and that it follows a gradient of functional interest along the dorso-ventral axis of layer II. Two features make this gene a potential candidate for understanding the nature of information processing throughout spatial navigation circuits. First, Arc is regulated by activity (Lyford et al., 1995) and can therefore be used as a marker of a cell's activity history during a task (Guzowski et al., 1999). Second, it has also been shown to be important for plasticity and homeostatic mechanisms (Plath et al., 2006; Shepherd et al., 2006), as well as for spatial learning (Plath et al., 2006). This is likely to be due to its interactions with structural synaptic components and AMPA receptors (Chowdhury et al., 2006). As such, novel-exploration induced Arc expression may be an effective indicator of learning-related neuronal activity in particular regions of the hippocampal-entorhinal circuit during learning, *in vivo*.

Arc expression can be up-regulated under a number of *in vitro* and *in vivo* conditions (see (Bramham et al., 2010) for review). It is not necessarily a direct marker of activity (Guzowski et al., 2006), but appears to be particularly modulated by novelty and the requirement for learning. Like *c-fos* and *zif268*, exposure to a novel environment such as a hole board (Ons et al., 2004), swimming (Gusev et al., 2005), open-field exploration (Guzowski et al., 1999), learning of water-maze tasks (Gusev et al., 2005), and stressful environments (Thompson et al., 2010) can all cause transcriptional up-regulation. Seizures also cause a particularly strong increase in *Arc* (Guzowski et al., 1999). This up-regulation is particularly interesting in CA1, where it is specific enough to enable the separation of recent events in time, based on whether the *Arc* RNA is present in the nucleus or cytoplasm of pyramidal cells (Guzowski et al., 1999). Exploration of a novel environment therefore seems an appropriate task for investigating a role for Arc in spatial navigation circuits.

Changes in *Arc* expression, resulting from experiences such as novelty exposure, have been observed across large areas of the brain, including much of the neocortex, particularly limbic, orbital and cingulate cortices, the hippocampus (Messaoudi et al., 2007), caudate putamen, lateral olfactory tract nucleus and basolateral amygdala complex (Ons et al., 2004). Despite this wealth of knowledge regarding *Arc* up-regulation, the precise spatial localisation of changes in *Arc* expression during simple behaviours such as active novel exploration have not been explored in MEC. There is some research that suggests that the learning of a water maze

task causes a strong increase in *Arc* RNA levels, whereas swimming alone produces only a moderate increase (Gusev et al., 2005) (see Fig 6.1 (i)).

Research considering the specificity of changes in *Arc* expression to particular layers, in MEC or elsewhere, is limited. In the work by Gusev et al. (2005) using rats in water maze tasks, the ISH data suggests that this up-regulation is strongest in the deep layers of MEC, the projection layer in which cells receive excitatory inputs from CA1 of the hippocampus and subiculum (Tamamaki and Nojyo, 1995), and send axons to other cortical areas (Insausti et al., 1997). However, these data were not quantified and images only show ventral MEC regions. If expression of Arc is higher in the deep layers than the superficial layers of MEC, this could indicate that Arc-dependent plasticity is differentially recruited across these regions. Differential expression of *Arc* in superficial and deep layers of posterior parietal cortex and granular insular cortex was recently addressed by Burke et al. (2005) and Takehara-Nishiuchi et al. (2012), who showed by looking at the precise localisation of changes in *Arc* RNA that superficial cortical neurons are more sensitive to spatial context than deep cortical neurons. It is not known if the superficial and deep layers of MEC, which contain different proportions of highly spatially tuned cells (Sargolini et al., 2006), show a similar pattern.

An important factor to consider when investigating the importance of Arc expression is that much of the research has studied only levels of RNA, when it is the protein that is directly necessary for the changes associated with plasticity. Arc transcription can be observed in the nucleus within minutes of stimulation and peaks in the cytoplasm after 30 minutes ((Guzowski et al., 2001), Fig. 6.1(II)). Arc protein can be translated locally in dendrites within very short time periods (Waung et al., 2008). Arc transcription and translation have different dependencies (see Bramham et al., 2010, for review). However, it is difficult to study these processes in parallel, or independently of one another. Mice that express an enhanced GFP construct under the control of the endogenous *Arc* promoter (Wang et al., 2006) provide a means of directly imaging *Arc*-GFP activation, a somatic indicator of cells expressing Arc, as well as Arc protein expression, using immunofluorescence. Additionally, homozygote GFP knockin mice can reveal *Arc* promoter activity in mice that do not express Arc protein. In visual cortex, GFP expression peaks at 2 hours following exposure to a novel environment (see Fig 6.1 (iii)), demonstrating similar results to the time course of *Arc* protein expression (Wang et al., 2006). *Arc* and GFP were co-expressed across many cells, but some cells uniquely expressed one or the other (see Fig. 6.1 (iv)). This suggests that Arc-GFP activation and Arc protein expression may not always directly parallel one another and may instead provide complementary information.

Taken together, it appears that investigating activity-dependent Arc-GFP activation and protein expression of *Arc* at the level of MEC layers could provide insight into the circuit organisation of MEC during activity. In this chapter, I outline how the expression of Arc changes when an animal is placed into a novel spatial environment, one in which cells across MEC, particularly in layer II, are known to be active (Hafting et al., 2005; Burgalossi et al., 2011).

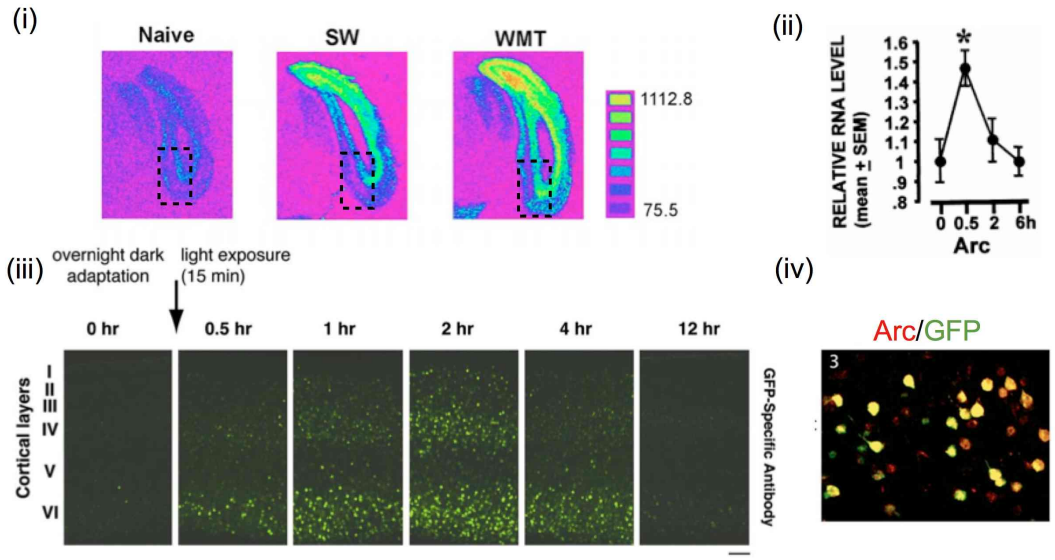


Figure 6.1: *Previous work exploring Arc up-regulation* (i) Colour-coded images showing radioactive *in situ hybridisation* for *Arc* mRNA in coronal sections containing the entorhinal area. The black square outlines the MEC region. RNA is higher following a water maze task (WMT) than following a swimming only condition (SW) or a naive condition (adapted from (Gusev et al., 2005)).(ii) Relative *Arc* mRNA expression is plotted as a function of time following spatial water task training. Immediate early gene RNA peaks at 30 minutes and returns to baseline by ~6 hours (adapted from (Guzowski et al., 2001)). (iii) Confocal images showing the time course of changes in GFP antibody fluorescence labelling in the visual cortex of an *Arc*-GFP knock-in mouse following exposure to a novel environment. GFP expression (GFP is under the control of the endogenous *Arc* promoter (Wang et al., 2006)) peaks at approximately 2 hours (from (Wang et al., 2006) Fig. 2a) (iv) Confocal image showing co-expression of *Arc* (red) and GFP (green) in an Arc^{+}/GFP^{+} mouse, labelled using antibodies (adapted from (Wang et al., 2006) Fig. 1c).

6.2.1 *Aims*

In this chapter I set out to explicitly characterise changes in *Arc* expression following novel exploration. I used heterozygous Arc^{+}/GFP^{+} mice, that express enhanced GFP under the control of the endogenous promoter to address several key questions:

1. Is *Arc* up-regulated in MEC following active exploration of a novel environment? This type of novel spatial experience should engage MEC and cause a strong effect on animals that have previously only experienced a cage environment.
2. Is there a gradient in *Arc* expression in MEC layer II following exposure to a novel environment, and is expression along the DV axis modulated differentially by exposure to a spatial environment?
3. How does *Arc* up-regulation compare across the different layers of MEC and with other regions of neocortex and hippocampus?

4. How do GFP expression and *Arc* protein expression compare? Arc-GFP mice are designed to report *Arc* promoter activity (Wang et al., 2006) so it is important to compare the expression patterns of Arc and GFP in the same mice.

5. Is functional Arc protein necessary for differential activation of *Arc*? I explore this using GFP^+/GFP^+ mice.

6.3 METHODS

In this chapter, I outline the second part of the experiment described in the previous chapter. To examine how exploration of a novel environment modulates the distribution of cells expressing *Arc*, I placed Arc^{+}/GFP^{+} mice into a square open-field arena containing novel objects for 15 minutes. Animals were perfused 90 minutes later to allow time for the translation of GFP and *Arc* protein. In these mice GFP is under the control of the *Arc* promoter, and could therefore be considered a reporter of *Arc* promoter activity (Wang et al., 2006), although the regulation of its translation could mask this (see 6.5.7). Briefly, I used two measures to quantify GFP labelling. 'Directly imaged' GFP was imaged directly from tissue sections. Adjacent sections were then co-labelled with either GFP antibody and a neuron-specific antibody, NeuN. I used one measure to quantify *Arc* protein, using adjacent sections to the GFP antibody, stained with *Arc* antibody and NeuroTrace, a fluorescent Nissl stain. Methods specific to this part of the experiment are outlined below. Methods relating to tissue preparation, immunofluorescence and image analysis are as described in the previous chapter.

6.3.1 *Procedure for exposure to a novel environment*

In order to analyse *Arc* expression following active up-regulation, I directly compared a 'novel-environment' mouse with a paired 'home-cage' control mouse (in most cases, a littermate). Both mice were removed from their home cage between 8.30 and 9am. While the 'Home-cage' mouse was immediately anaesthetised and perfused, the 'novel-environment' mouse was placed into a novel arena containing 8 novel objects for 15 minutes, then returned to an empty cage in a quiet room for 90 minutes prior to perfusion. Previous work has shown that GFP levels peak at 1-3 hours (Wang et al., 2006) while Arc protein levels, measured using Western blot or immunohistochemistry, peak at 2-4 hours after *in vitro* stimulation (Wallace et al., 1998; Wang et al., 2006; Messaoudi et al., 2007; Rodríguez et al., 2005; Lyford et al., 1995). I therefore chose this time period to capture activity-dependent protein expression for GFP and Arc whilst minimising the time in which other experiences could modify protein expression. Brains were processed identically from this point (see Fig 6.2(A)). The arena was 80 x 80cm with 40cm high walls and contained 8 novel objects with interesting textures located in a circle around the arena (see Fig 6.2(B)). Visual patterns were attached to the walls of the arena. All mice were observed while in the arena using a camera mounted directly above that was connected to a computer and monitor. All mice showed considerable exploration of the arena, including attention to and climbing on novel objects.

6.3.2 *Quantification*

In the previous chapter I described several ways of measuring the GFP antibody, direct GFP and *Arc* antibody expression. These included ROI-wide measures of pixel intensity, which provides a coarse estimate, and the soma-surround measure, which provides a cell-based measure. In this chapter, I have chosen to present data using only the soma-surround measure because this minimises differences in background intensity across regions and mice. I have

not used ROI-wide measures because they show high variability and are subject to the most confounding variables.

6.3.2.1 *Normalisation*

I have also included a normalisation step after calculation of the soma-surround. These data were collected from mice across 4 sessions (3 for Arc^+/GFP^+, 2 for GFP^+/GFP^+) and despite efforts to minimise differences across sessions, it is possible that variability in immunofluorescence, laser power and the conditions of the mice prior to sacrifice could affect overall measurements and therefore compromise the ability to detect differences between home-cage and novel-environment animals.

To normalise the data across sessions, I adjusted all data for each ROI by the session mean for that ROI. The exception to this was in MEC layer II where all regions were normalised together. First, I calculated the mean soma-surround measure for each mouse within each condition within each session. I then averaged the data across mice within a condition, then across conditions within a session. Penultimately, I divided every data point for each mouse by the mean for the session for that particular ROI (Fig. 6.2(C)).

I added one further step to simplify interpretation of the final results. All normalised data points were standardised to the approximate range [0,1] by using the 1% and 99% percentiles. I chose these percentiles instead of using the minimum and maximum values, so as to minimise the likelihood of extreme anomalies skewing the adjusted distributions. All measures are presented in histograms in the range [0,1], but all measurements for summary data were calculated using the full data range. As such, a soma-surround score of 0.5 represents a score that is half way through the range of scores observed.

6.3.2.2 *Thresholds*

In the previous chapter I compared regions based on the proportion of neurons or detected cells that exceed a threshold level of protein expression. For this chapter I have again calculated these thresholds. For GFP the 'low' threshold is calculated using a value defined using the distribution of scores for control tissue, such that 'low' cells express a detectable amount of protein that is unlikely to be due to chance. This value is the same across all regions because I only measured control expression in visual cortex. The moderate threshold indicates that a cell expresses an amount of protein in the upper half of the range of scores detected for a particular ROI (> 0.5 normalised range). The high threshold indicates that the score is in the top 10% of scores for GFP antibody labelling, and the top 25% for direct GFP labelling, as cell numbers are lower and more variable.

For Arc, I used a 'low' threshold calculated using the distribution of control data compared with that of non-control visual cortex data across both groups. Cells with a soma-surround score exceeding 0.16 of the normalised range are likely to be true Arc-expressing cells. The moderate and high thresholds were calculated as the upper 50% and 25% of the range, respectively. These indicate the number of cells in each group that express particularly high levels of Arc.

For direct GFP fluorescence for GFP^+/GFP^+ mice, the histograms were very skewed so I adjusted the thresholds to 25% and 50% to provide more information about the distribution

6.3.3 *Statistical methods*

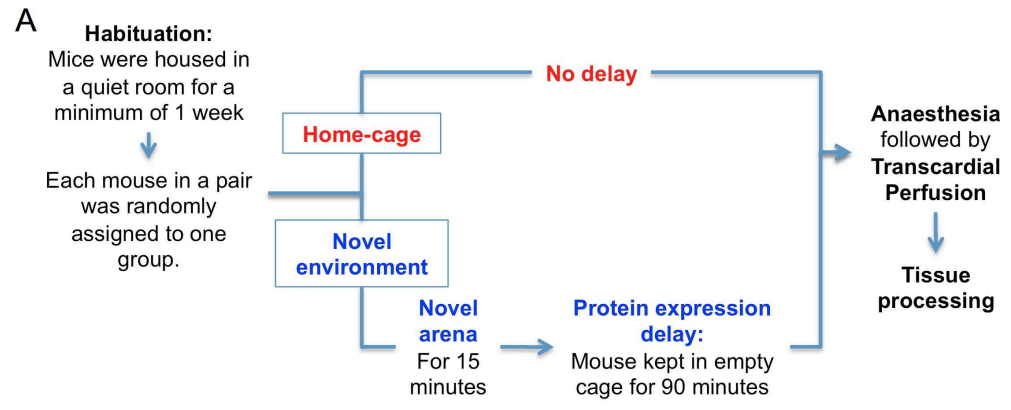
In this chapter my primary aim was to consider the effect of novel exploration on individual brain regions. I have therefore chosen to present the results of each region separately and use one-way independent-measures ANOVAs (equivalent to a student's t-test) to test the effect of condition (home-cage vs. novel-environment) on Arc expression.

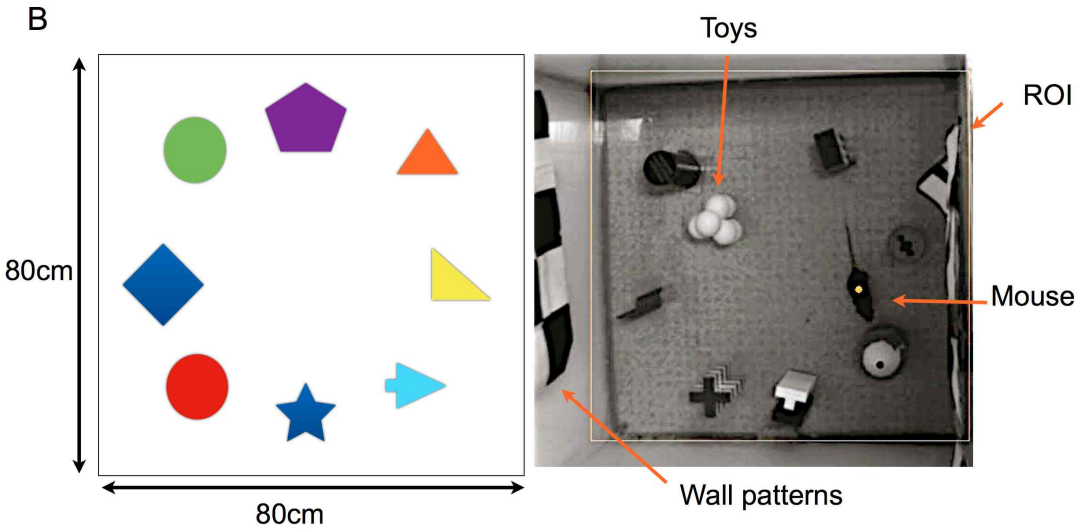
I did not include Region as a second factor and use a two-way ANOVA because I did not want to distract from the main question of the effect of condition. I also chose not to use mouse genotype as a third factor on the basis that this was considered an independent experiment. Genotype remains an important question, but the data was not collected for both genotypes simultaneously, and therefore it is difficult to interpret comparisons. None of the pairs included a mixture of GFP^+/GFP^+ and Arc^+/GFP^+ mice.

For all statistical comparisons, to minimise the possibility of obtaining significant effects due to chance, I have presented the results of a multivariate ANOVA (MANOVA) to test the effect on all measures simultaneously. This test is not always appropriate; in some of the comparisons, this test did not necessarily meet the required assumptions, partially due to the low number of samples. Furthermore, for simplicity I used the same measures across all comparisons, but because the distributions of data vary, some measures are not always appropriate for inclusion in the MANOVA. I have therefore presented the results of further ANOVA tests on each dependent variable both where the MANOVA test was significant, and where it was not significant ($\alpha t \alpha < 0.05$), but the results of at least two of the subsequent ANOVA tests were.

To test for differences in gradient between conditions, I used the analysis of covariance (ANCOVA) test, which combines ANOVA with regression.

For the majority of comparisons, data have been plotted using box plot diagrams overlaying the actual data points (grey circles). The mean for each condition is indicated by a red circle (home-cage), blue diamond (novel-environment), green triangle (home-cage KO), or black square (novel-environment KO).





C

Sessions	1						2						3					
Group	HC			NE			HC			NE			HC			NE		
Mice	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r
Group mean	mean(abc)			mean(def)			mean(ghi)			mean(jkl)			mean(mno)			mean(pqr)		
Session mean	s1 = mean(mean (abc),mean(def))						s2 = mean(mean (ghi),mean(jkl))						s3 = mean(mean (mno),mean(pqr))					
Normalise	a/ s1	b/ s1	c/ s1	d/ s1	e/ s1	f/ s1	g/ s2	h/ s2	i/ s2	j/ s2	k/ s2	l/ s2	m/ s3	n/ s3	o/ s3	p/ s3	q/ s3	r/ s3

Figure 6.2: *Procedure for 'home-cage' and 'novel-environment' animals* (A) A schematic showing the behavioural procedures for the two groups of mice (home-cage and novel-environment). Novel-environment animals were placed into a square arena for 15 minutes, followed by 90 minutes in an empty cage, whereas home-cage mice were directly removed from their cage. (B) Schematic and photo showing the 80 x 80 x 40 (height) cm novel environment used. Eight unique objects, either made from lego or just hard plastic, were placed in a circle within the arena to encourage exploration. Patterned images were attached to the walls. The novel environment was placed on the floor of a well-lit room with a camera above it to enable monitoring and recording of animal movements throughout the exploration period. (C) Schematic showing normalisation procedure for example mice. For each group and each session the mean soma-surround score is calculated. All data from each mouse are then divided by the mean score corresponding to the session that the mouse was in.

6.4 RESULTS

6.4.1 *In superficial MEC, exposure to a novel environment does not cause up-regulation of GFP*

In the previous chapter I demonstrated that *Arc* -GFP activation is low at baseline across different regions in MEC. In this section, I focus on two questions: is Arc-GFP activation increased following novel exploration, and is GFP expression at different levels along the dorso-ventral axis of MEC following novel exploration?

6.4.1.1 *GFP expression is indistinguishable from baseline following novel exploration*

As shown in Figure 6.3, GFP labelling is low in MEC layer II across almost all animals following novel exploration. GFP^+ cells can be detected by eye in just 3 of the animals, and only in the more dorsal half of MEC. The somatic expression of these cells is very similar to that for home-cage animals, as shown by the highly overlapping distributions for the most dorsal region (Fig. 6.4 (A)). A MANOVA test measuring the effect on 3 dependent variables of GFP expression reveals no effect of condition (Fig. 6.4 (B), $F_{(1,34)} = 1.38, p = 0.262, n = 8, 8$), and average images show no clear differences (Fig. 6.4 (C)).

Directly imaged somatic GFP labelling in the two conditions show no significant differences in the total number of GFP^+ cells (Fig. 6.4 (D)(a), ANCOVA: $F_{(1,98)} = 1.064, p = 0.304$) or the number of moderately or highly GFP-labelled cells (Fig.6.4 (D)(b-c), ANCOVA Mod: $F_{(1,98)} = 1.20, p = 0.275$, High: $F_{(1,98)} = 0.862, p = 0.355$). Differences in mean soma-surround were not tested due to the low number of detected cells in each region.

GFP^+ neurons can also be identified using GFP antibody and NeuN labelling overlaid (Figure 6.5). GFP labelling within neurons is relatively low. Quantification of this data reveals that novel-environment animals show very similar somatic expression to home-cage animals both in the most dorsal region (Figs. 6.6 (A)) and throughout the entire MEC extent (approximately 1.5mm from the dorsal border) (Figs. 6.6 (C)). There is no overall effect of group ($F_{(1,107)} = 1.64, p = 0.155, n = (8, 8)$ when using a MANOVA to test the mean and variability in soma-surround scores and the proportion of low-, mid- and high-GFP expressing neurons together. There is also no effect on individual measures (Figs. 6.6 (B-C), ANCOVA Mean : $F_{(1,114)} = 0.924, p = 0.338$, SD : $F_{(1,114)} = 2.498, p = 0.116$, Mod : $F_{(1,114)} = 0.036, p = 0.849$, High : $F_{(1,114)} = 1.442, p = 0.232$), except for the proportion of cells exceeding a low threshold (Figs. 6.6 (C)(c), Low : $F_{(1,114)} = 5.215, p = 0.024$), which is actually higher in the home-cage group.

6.4.1.2 *GFP expression following exposure to a novel environment shows a weak dorso-ventral gradient*

In the previous chapter I showed that under baseline conditions, GFP expression is higher in dorsal areas of MEC layer II than ventral regions, but that the gradient is quite shallow. In novel-environment animals the cell number detected directly using GFP fluorescence was too few for analysis (Fig. 6.4(D)(a)), so I did not directly question the presence of a gradient, particularly as cell density may decrease along this axis, thus confounding the results of cell counts (Section 5.4.1.3).

Identifying visible gradients in expression in each mouse using the GFP antibody is also difficult due to low expression levels (Figure 6.5). However, the use of ANCOVA reveals significant effects of region across home-cage and novel-environment animals for the mean and variability in GFP soma-surround score across neurons (Fig. 6.6 (C)(a-b), Mean : $F_{(7,114)} = 2.10, p = 0.049$, SD : $F_{(1,114)} = 5.55, p = 0.00002$). There is no interaction between group and region (Fig. 6.6 (B), MANOVA: $F_{(1,107)} = 0.597, p = 0.967$, ANCOVA $p > 0.7$). The proportion of neurons that exceed thresholds also varies significantly (Fig. 6.6 (C)(c-e)), with 20% of the most dorsal neurons exceeding a moderate threshold, compared with less than 5% of the most ventral neurons across both groups. This suggests that dorsal neurons have an increased propensity to express moderate levels of Arc, but a similar proportion of neurons express Arc at low levels (Fig. 6.6 (C)(c)).

To determine whether animals demonstrate a gradient after exploration, I performed a linear regression analysis using the novel-environment data, normalised for each mouse. The mean soma-surround significantly decreases along the gradient, but the slope is extremely shallow; on average, the somatic region is just half a unit of pixel intensity less greater than the surround in ventral cells than in dorsal cells (Fig. 6.6 (E)(a), slope = -0.4, r^2 = 0.07, p = 0.045). There is a more significant gradient in the variability between cells (Fig. 6.6 (E)(b),slope = -0.58, r^2 = 0.17, p = 0.0008), but this is also shallow. The gradient for the proportion of neurons exceeding a moderate threshold is not significant (Fig. 6.6 (E)(c), p = 0.064).

In summary, exploration of a novel environment does not up-regulate GFP expression and it does not distinguish GFP expression in dorsal and ventral regions.

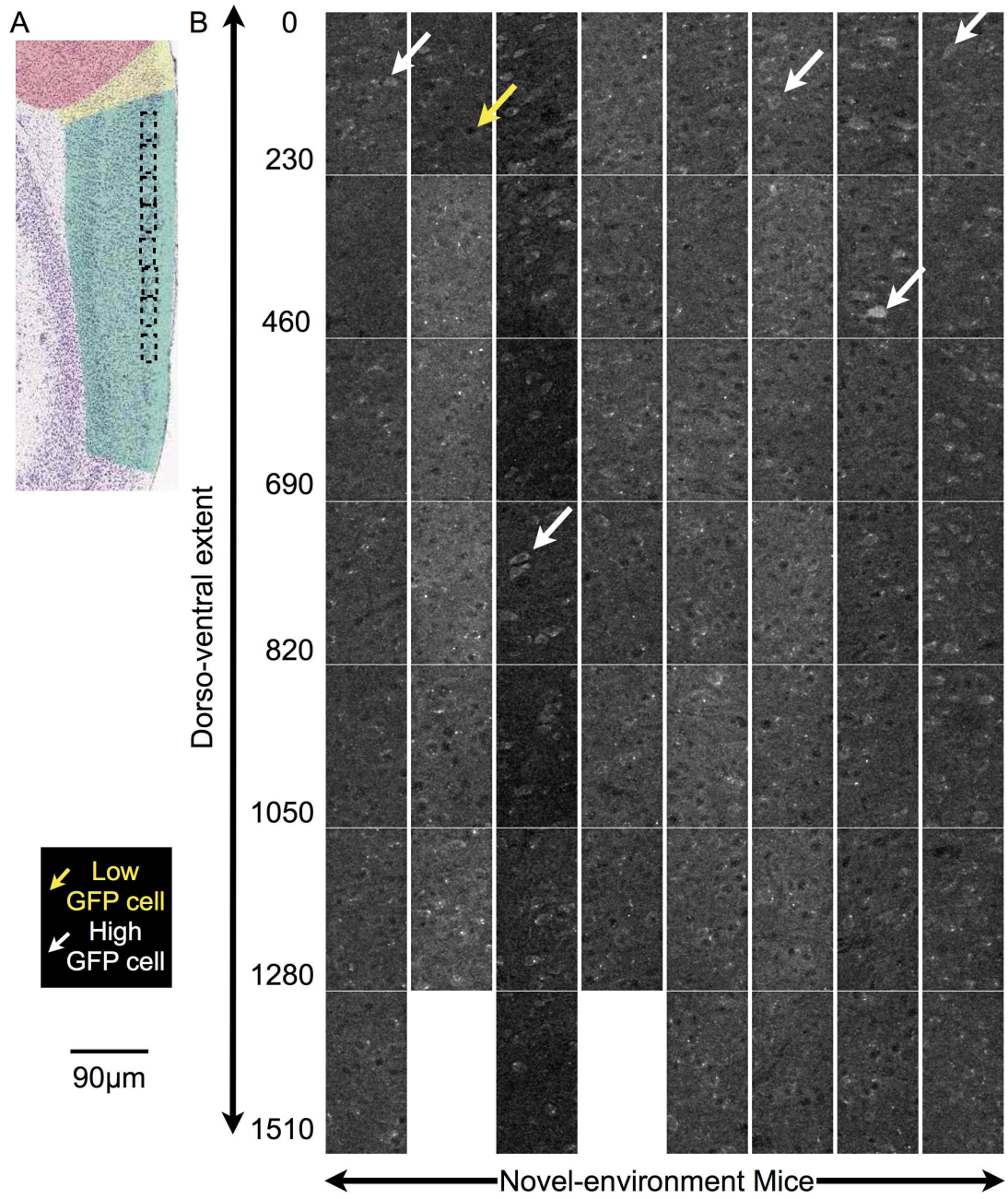


Figure 6.3: *Directly imaged GFP fluorescence along the dorso-ventral axis of MEC layer II in mice that have been exposed to a novel environment.* (A) Nissl-stained sagittal section image of the MEC region (colour-coded green) to show the approximate location of the regions shown in (B). (B) Confocal 20x maximum projection images showing direct GFP fluorescence across all ROIs of MEC layer II from all novel-environment animals (n=8). Scale (left) shows the approximate distance of the section from the dorsal border of MEC with the large patches (in µm).

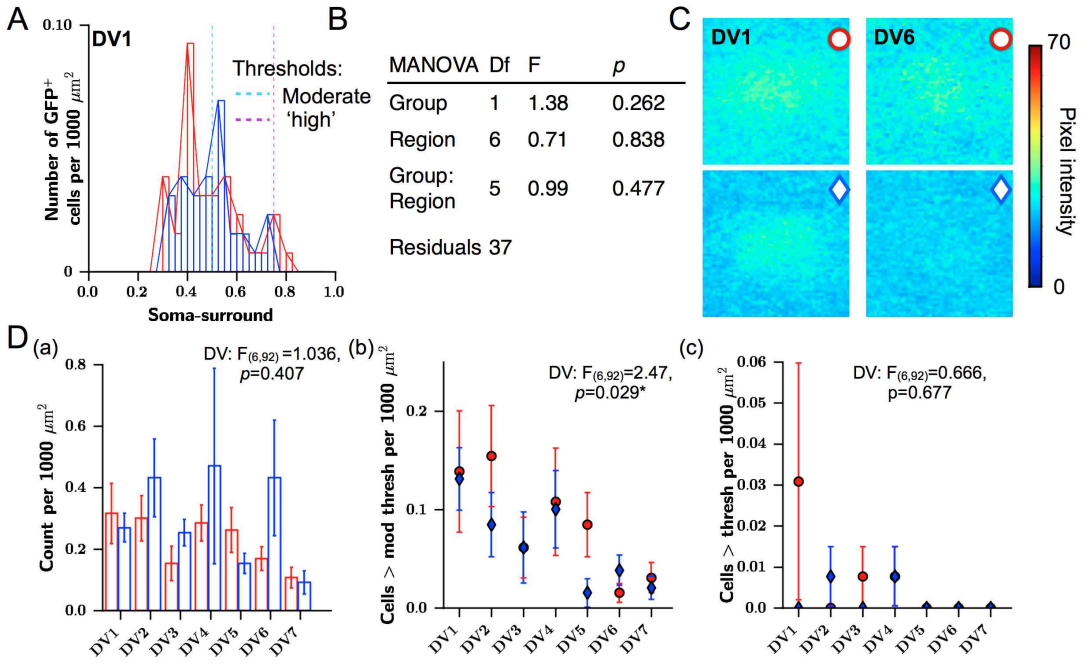


Figure 6.4: *Quantifying the effect of exposure to a novel environment on direct GFP fluorescence along the DV gradient in MEC layer II.* (A) Histograms show the overlap between the distributions of cells detected in the most dorsal MEC layer II region for home-cage (red) and novel-environment (blue) mice. Thresholds are indicated. (B) Table showing that a MANOVA test indicates no effect of group on GFP expression. (C) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice in the most dorsal region (DV1) and a more ventral region (DV6). (D) Soma-surround measures for cells detected directly from GFP fluorescence: (a) Number of cells detected per 1000μm^2, (b-c) Number of cells detected that exceed the (b) moderate and (c) high soma-surround thresholds. Results for effect of region using a two-way ANCOVA are shown (D), with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

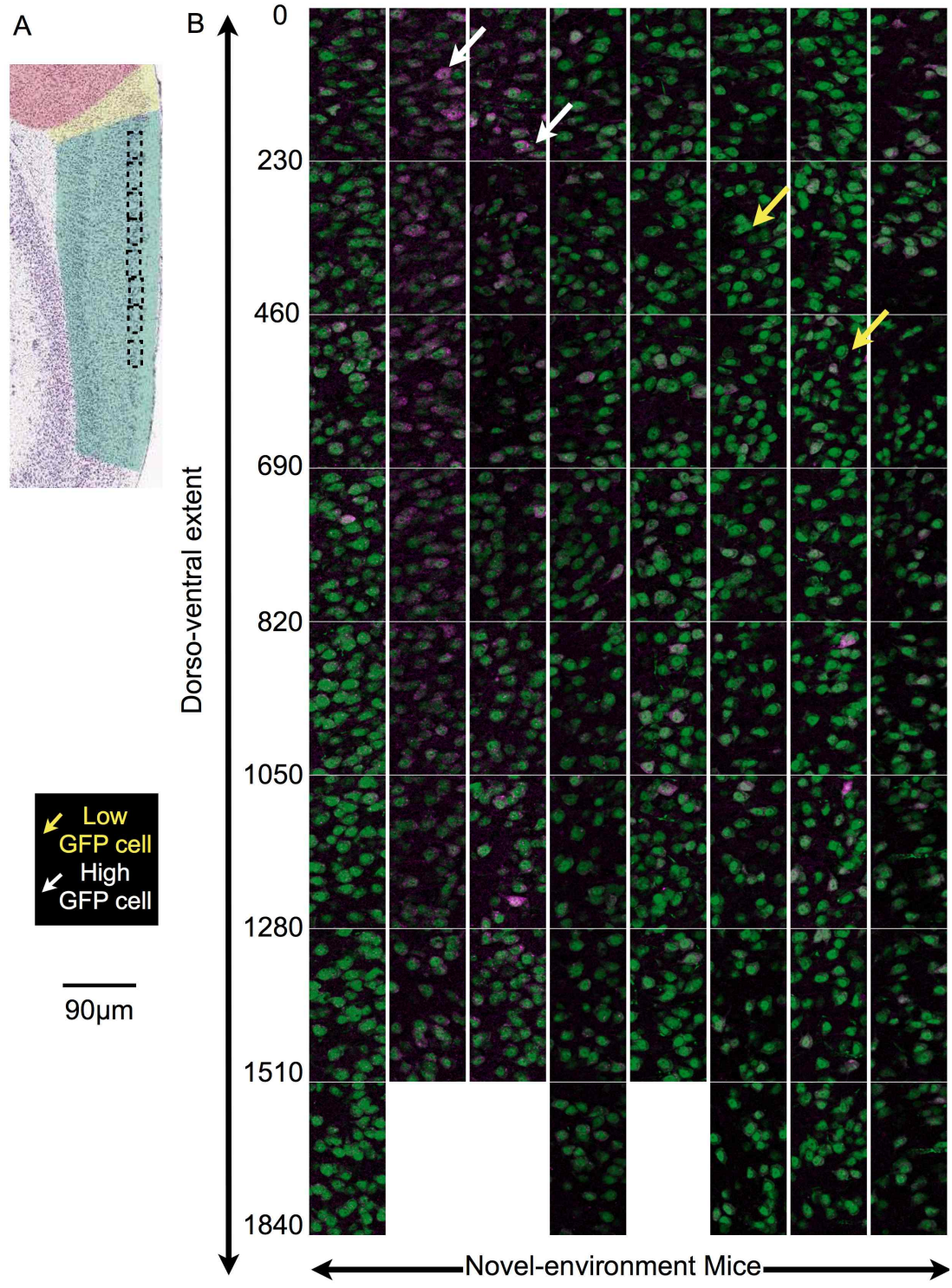


Figure 6.5: *GFP antibody labelling along the dorso-ventral axis of MEC layer II in mice that have been exposed to a novel environment.* (A) Nissl-stained sagittal section image of the MEC region (colour-coded green) to show the approximate location of the regions shown in (B). (B)Confocal 20x maximum projection images showing the overlay of GFP antibody fluorescence (magenta) and NeuN labelling (green) across all ROIs of MEC layer II from all novel-environment animals (n=8). Scale (left) shows the approximate distance of the section from the dorsal border of MEC with parasubicular patches (in µm).

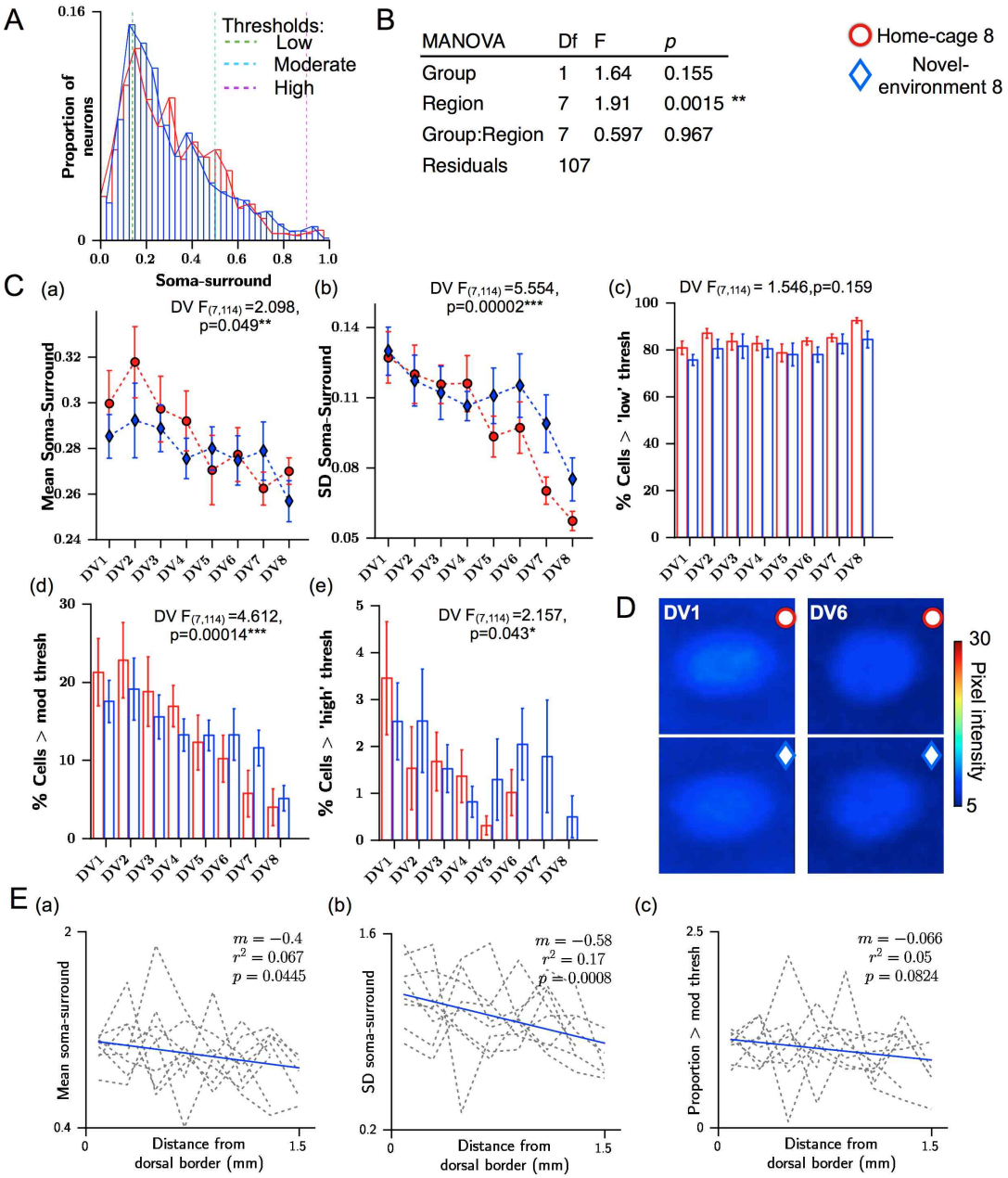


Figure 6.6: *Quantifying the effect of exposure to a novel environment on GFP expression along the DV gradient in MEC layer II.* (A) Histograms show the overlap between the distributions of soma-surround scores for all neurons detected in the most dorsal MEC layer II region for home-cage (red) and novel-environment (blue) mice. Thresholds are indicated. (B) Table showing that a MANOVA test indicates no effect of group on GFP expression, but an effect of region. (C) Summary Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean soma-surround score (b) the standard deviation in soma-surround score, (c,d,e) The percentage of neurons expressing (a) a 'low', (b) a 'moderate', (c) a 'high' threshold level of GFP compared to the surround. (D) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron, calculated across all detected cells across all mice in the most dorsal region (DV1) and a more ventral region (DV6). Results for effect of region using a two-way ANCOVA are shown (D), with post-hoc significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. (E) Linear regression plots showing the relationship between dorso-ventral extent and (a) mean soma-surround score, (b) standard deviation in soma-surround score, and (c) proportion of neurons exceeding a moderate threshold. Blue line shows line of best fit, where m is the gradient, and r^2 is calculated using data normalised by the individual mean for each animal. Grey lines show scores for each novel-environment mouse.

6.4.1.3 *Novel exploration selectively increases Arc protein expression in MEC layer II*

In MEC layer II, exposure to a novel environment appears to have no effect on Arc-GFP activation. I next asked if novel exploration affects the expression of Arc protein. The images in Fig. 6.7 (A) show that in both dorsal and ventral regions for this pair of mice, novel exploration appears to modestly increase *Arc* protein labelling across the ROI. It is possible to notice in these images that although Arc labelling often coincides with NeuroTrace labelling, in MEC layer II cells it shows a particular tendency to lie around the border of the defined cell, presumably due to its localisation to synapses. Also, there are many brightly NeuroTrace-labelled cells that do not resemble NeuN-labelled neurons and are likely to be glial cells. As such, I decided to identify cells directly using Arc protein labelling (blind to condition). I defined the soma and surround regions using similar criteria to directly imaged cells.

Histograms of all the manually detected cells in dorsal and ventral regions show that more cells can be detected following exposure to a novel environment (Fig. 6.7 (B)(i)). The mean somatic Arc protein labelling (compared to surround) of cells is significantly higher in layer II of novel-environment (ne) than home-cage (hc) animals (Fig. 6.7 (C)(a)): $F_{(1,12)} = 15.32, p = 0.002, n = (hc : 3, ne : 4)$), as is the number of cells that express moderate or high levels of Arc protein (Fig. 6.7 (C)(b)): Mod: $F_{(1,12)} = 11.0, p = 0.0062, High : F_{(1,12)} = 8.14, p = 0.015, n = (hc : 3, ne : 4)$). The number of detectable cells does not significantly increase ($F_{(1,12)} = 4.464, p = 0.0562$). These findings are only significant for dorsal regions, but there is no interaction between condition and region (Mean: $F_{(1,12)} = 0.318, p = 0.583$).

Dorsal regions show a tendency towards higher mean cellular expression ($Mean : F_{(1,12)} = 4.64, p = 0.052$) and increased number of high-expressing neurons ($F_{(1,12)} = 42.90, p = 0.114$), but there is no robust difference between dorsal and ventral regions.

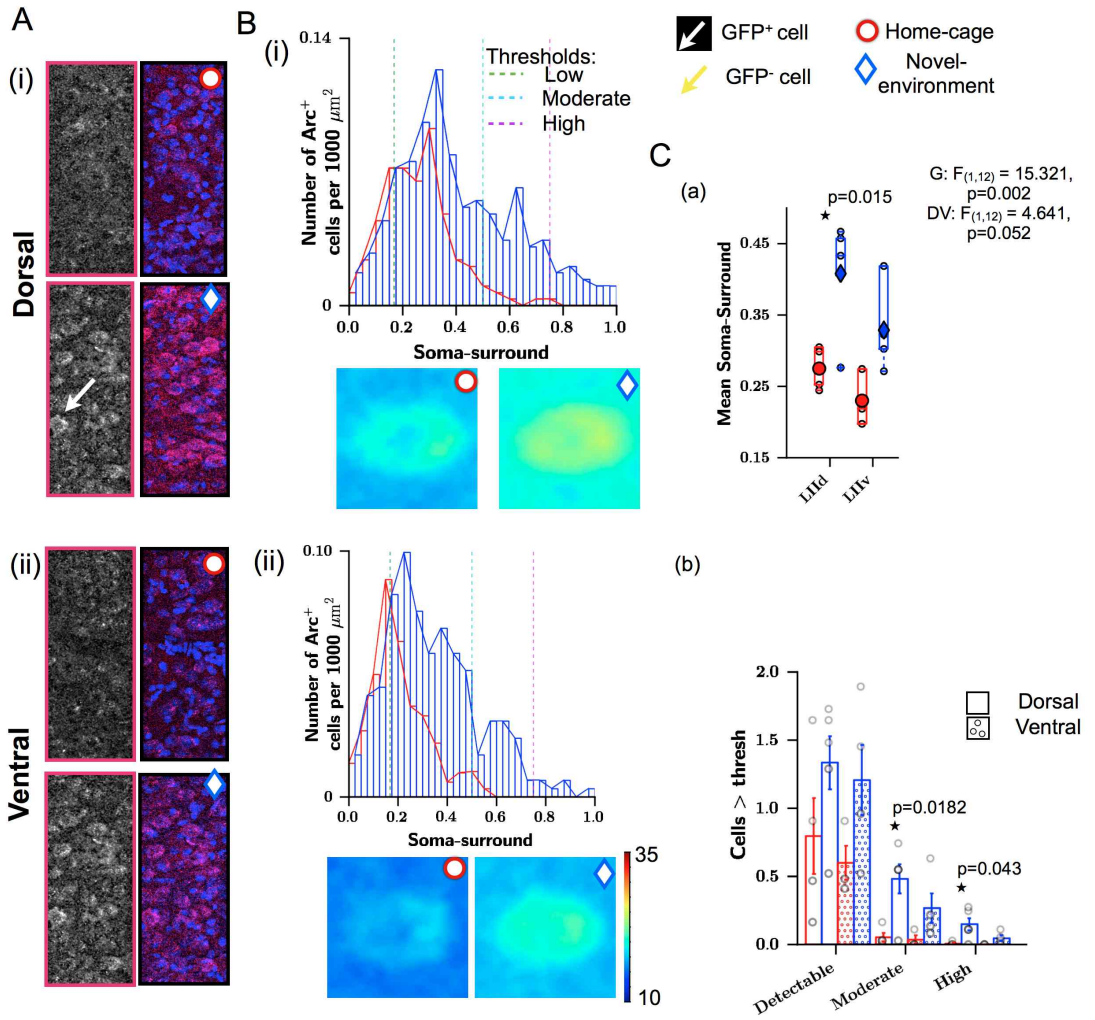


Figure 6.7: *Novel exploration increases Arc protein expression in MEC layer II* (A) Example confocal 20x maximum projection images of (i) dorsal MEC layer II and (ii) ventral MEC layer II showing the difference in Arc protein labelling (magenta outline) for a representative pair of home-cage (red circle, upper) and novel-environment (blue diamond, lower) mice. Arc protein was co-labelled with NeuroTrace (blue). (B) Histograms show the distributions of Arc soma-surround labelling fluorescence scores for all manually detected cells in (i) dorsal layer II and (ii) ventral layer II. Images show the average 60x60 pixel region of Arc antibody labelling surrounding a detected cell, calculated across all detected cells across all mice in each group. (C) Soma-surround measures for Arc antibody labelling within detected cells: (a) Mean soma-surround score, (b) the number of cells (per 1000 μm^2) with a soma-surround score exceeding a 'low', 'moderate' or 'high' threshold. Asterisks indicate significant results after a two-tailed t test on the effect of condition at alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$, n = (3,4).

6.4.2 *In MEC deep layers, exposure to a novel environment causes up-regulation of Arc*

The deep layers are thought to retrieve the major hippocampal projection into the MEC (Swanson et al., 1978) and provide the major output from the MEC to other cortical regions (Insausti et al., 1997). I therefore investigated whether exploration of a novel environment, which would be expected to activate hippocampal place fields, alters Arc expression in MEC deep layers.

6.4.2.1 *GFP expression increases following novel exploration*

In the previous chapter I demonstrated that *Arc* -GFP activation is very low at baseline in MEC deep layers. The images shown in Fig.6.8(A) show that following exploration of a novel environment the number of directly detectable GFP^+ cells in MEC deep layers is dramatically increased. This is particularly evident in the group histograms, which emphasise an increased number of detectable cells showing a range of GFP labelling (Fig.6.8 (B)). The group data shows that although novel exploration does not significantly increase the total number of cells detected, it increases the number of cells expressing a moderate level of GFP (Fig. 6.8 (C)(a)): $p = 0.03, n = (8, 8)$) and the mean GFP soma-surround labelling of detected cells (Fig. 6.8 (C)(b)): $p = 0.0016, n = (8, 8)$).

Data for somatic GFP antibody labelling within neurons also show a clear increase following novel exploration (Fig.6.8 (E)). GFP antibody labelling within NeuN-defined neurons provides estimates of the proportion of neurons that show increased expression following novel exploration. The shape of the distribution of soma-surround scores changes, indicating that both average Arc expression within cells and the variability between cells is increased (Fig.6.8 (F-G)). The mean GFP labelling within the soma of neurons relative to their surround increases significantly following novel exploration (Fig. 6.8 (H)(a)): $p = 0.0008, n = (8, 8)$). The variability across neurons also increases (Fig. 6.8 (H)(b)): $p < 0.0001, n = (8, 8)$). It follows that the proportion of moderately GFP labelled cells increases, from under 10% to over 20% of cells (Fig. 6.8 (F)(c)): $p = 0.0002, n = (8, 8)$) and the proportion of highly GFP labelled cells increases from none to just under 5% (Fig. 6.8 (F)(d): $p < 0.0001, n = (8, 8)$).

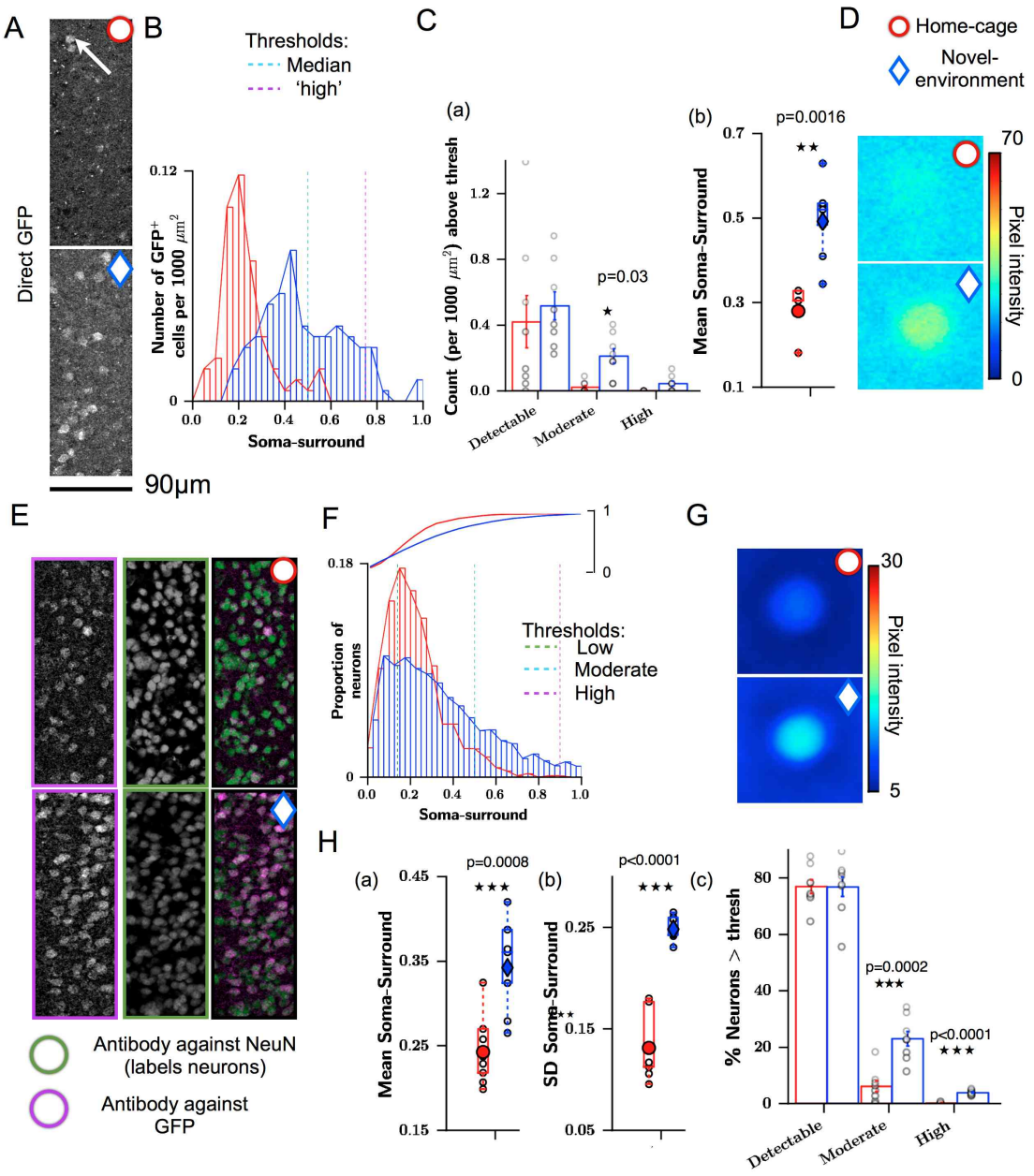


Figure 6.8: *Novel exploration increases GFP expression in MEC deep layers.* (A) Example confocal 20x max. projection images of direct GFP fluorescence in dorsal MEC deep layers for a home-cage (red) and a novel-environment (blue) mouse. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 μm^2 . (C) Direct GFP summary measures: (a) The number, in total and exceeding a moderate or high threshold of GFP^+ cells per 1000μm^2 is summarised across all mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all home-cage and novel-environment animals. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) the standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron (using the corresponding NeuN labelling), calculated across all detected cells across all mice in each group. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.4.2.2 *Arc protein expression increases following novel exploration*

The data presented above suggest that *Arc* -GFP activation is selectively up-regulated in MEC deep layers. To test whether *Arc* protein expression follows this pattern, I performed immunofluorescence with an anti-*Arc* antibody and quantified expression directly. Sections were co-stained with NeuroTrace for visualisation purposes.

As discussed in the previous chapter (5.4.2) and shown again in Fig. 6.9(A), under baseline conditions it is difficult to visually detect any cells in MEC deep layers that are labelled for *Arc* protein. The images in Fig. 6.9(A) demonstrate that exploration of a novel environment dramatically increases the detectability of Arc-labelled cells in MEC deep layers, but it is evident that only a subset of cells labelled with NeuroTrace are labelled for *Arc* protein.

Across animals, novelty increases the number of cells expressing low, moderate and high levels of *Arc* protein (Fig. 6.9 (B)(C)(a): $\text{Low} : p = 0.0027, \text{Mod} : p = 0.025; \text{High} : p = 0.044, n = (5, 7)$ and the level of *Arc* protein antibody labelling within cell somata relative to their surround (Fig. 6.9 (B),(C)(a): $p = 0.042, n = (5, 7)$). This increase is evident in the images of average cells (Fig. 6.9 (D)).

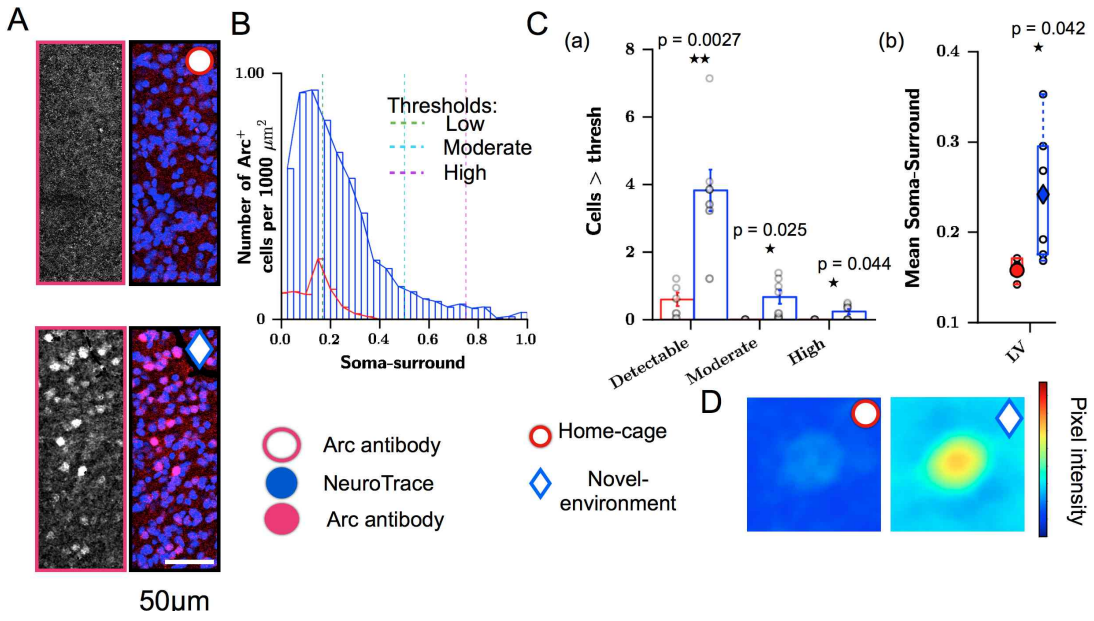


Figure 6.9: *Novel exploration increases Arc protein expression in MEC deep layers.* (A) Example confocal 20x maximum projection images showing the difference in Arc protein labelling (magenta outline) for a representative pair of home-cage (red circle, upper) and novel-environment (blue diamond, lower) mice. Arc protein was co-labelled with NeuroTrace (blue). (B) Histograms show the distributions of Arc soma-surround labelling fluorescence scores for all detected cells. (C) Soma-surround measures for Arc antibody labelling within detected cells: (a) the number of cells with a soma-surround score exceeding a low, 'moderate' and 'high' threshold (per 1000 μm^2) (b) the mean soma-surround score. (D) Images show the average 60x60 pixel region of Arc antibody labelling surrounding a detected cell, calculated across all detected cells across all home-cage (red,*n*=5) and novel-environment (blue, *n*=7). Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.4.3 *Novel exploration does not cause a detectable increase in GFP expression in the large patches dorsal to MEC.*

The number of GFP^{+ve} cells that can be detected using direct GFP fluorescence is relatively low in the large putatively parasubicular patches in home-cage and novel-environment animals (Fig. 6.10(A)). There is considerable variability across animals, particularly in the home-cage group, in which a larger number of cells overall, can be detected (Figs. 6.10(B) and (C)(a)). A MANOVA test across the four dependent variables reveals no effect of condition ($F_{(1,14)} = 1.016, p = 0.44, n = (8,8)$). These variables include the number of cells detected overall and exceeding certain thresholds of soma-surround score (Fig. 6.10 (C)(a): ANOVA -Total: $F_{(1,14)} = 1.791, p = 0.20$, Mod: $F_{(1,14)} = 0.962, p = 0.343$, High: $F_{(1,14)} = 1.615, p = 0.224$), and the mean soma-surround of these detected cells (Fig. 6.10 (C)(b), ANOVA: $F_{(1,14)} = 0.235, p = 0.636$). Averaging cellular expression across the two groups highlights their similarity (Fig. 6.10 (D)).

GFP labelling of neurons appears to be similar in home-cage and novel-environment animals (Fig. 6.10(E)) and the distribution of soma-surround scores across all neurons, averaged across mice, is similar in the two groups (Fig. 6.10(F)). It follows that average somatic GFP labelling across all neurons is also similar (Fig. 6.10(H)). A MANOVA test reveals a significant effect overall on the 5 dependent variables measures ($F_{(1,14)} = 4.11, p = 0.0274, n = (8,8)$), but on no measure individually. Novelty exposure does not affect mean soma-surround GFP labelling of neurons (Fig. 6.10 (G)(a): $F_{(1,14)} = 2.304, p = 0.15, n = (8,8)$), the variability across neurons (Fig. 6.10 (G)(b): $F_{(1,14)} = 0.008, p = 0.93, n = (8,8)$), or the proportion of neurons exceeding a low, moderate or high level of GFP (Fig. 6.10 (G)(c): Low: $F_{(1,14)} = 4.518, p = 0.052, Mod : F_{(1,14)} = 2.099, p = 0.17, High : F_{(1,14)} = 0.228, p = 0.64, n = (8,8)$).

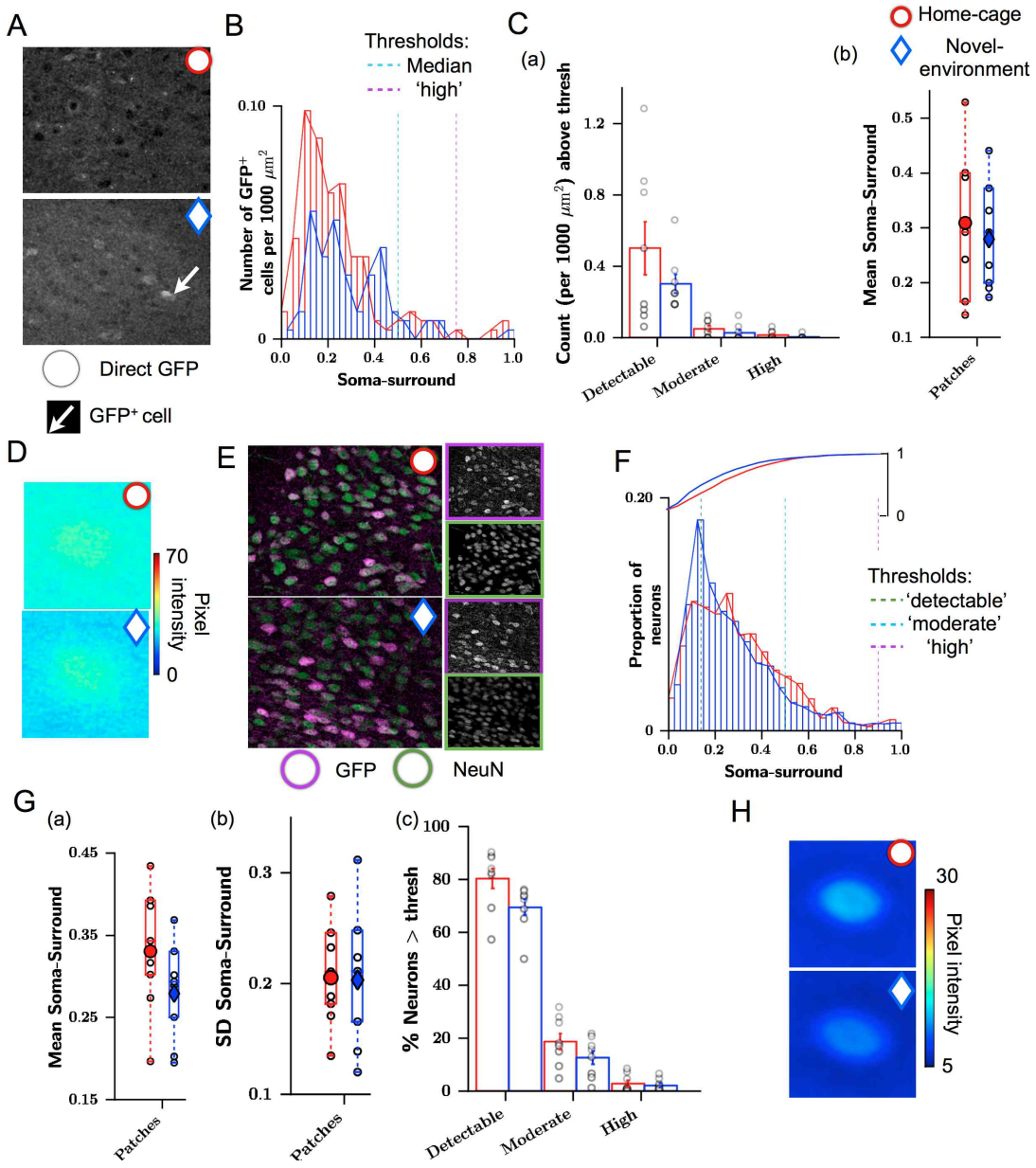


Figure 6.10: *GFP expression in the large patches is not significantly modulated by novel environment exposure* (A) Example confocal 20x max. projection images of dorsal patches showing direct GFP fluorescence for a representative pair of mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 μm^2 . (C) Direct GFP summary measures: (a) The number, in total and exceeding a moderate and high threshold, of GFP^+ cells per 1000 μm^2 is summarised across all home-cage (red,*n=8*) and novel-environment (blue, *n=8*) mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all home-cage and novel-environment animals. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron (using the corresponding NeuN labelling), calculated across all detected cells across all home-cage and novel-environment animals. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.4.4 *In the hippocampus, novelty does not cause detectable increases in GFP expression*

I have thus far shown that exploration of a novel environment does not change Arc-GFP activation in the MEC superficial layers or dorsal patches, but causes a clear increase in the deep layers of MEC. Previous work has shown that the dentate gyrus and CA1 regions of the hippocampus, which receive and send projections to these MEC layers, show increased transcription following maximum electroconvulsive shock treatment and exploration of novel environments (Guzowski et al., 1999; Messaoudi et al., 2007). I therefore examined the effects of novel exploration in these regions.

In CA1, under baseline conditions, very few GFP^+ cells can be detected in images showing the pyramidal layer, but this does not appear to increase in novel-environment mice (Fig.6.11(A)). The distributions of soma/CA1 non-somatic region score for detected GFP^+ cells show clear overlap between the two groups (Fig.6.11(B)). Using a MANOVA I show that there are no differences between the groups overall ($F_{(1,14)} = 0.44, p = 0.78, n = (8,8)$) including the number of detectable cells at different thresholds (Fig. 6.11 (C)(a)) or in the mean soma-surround score for these cells (Fig. 6.11 (C)(b)). Average neuronal images for the two groups are shown in Fig.6.11 (D).

These results are very similar for somatic GFP labelling of neurons. Images show that novelty does not increase the number of neurons expressing moderate levels of GFP (Fig.6.11(E)), which is confirmed across the groups in the overlapping histograms (Fig.6.11(F)). Mean somatic expression, variability in somatic expression and the proportion of neurons exceeding thresholds do not differ overall between the groups (MANOVA,$F_{(1,14)} = 1.05, p = 0.44, n = (8,8)$; Fig. 6.11 (G)).

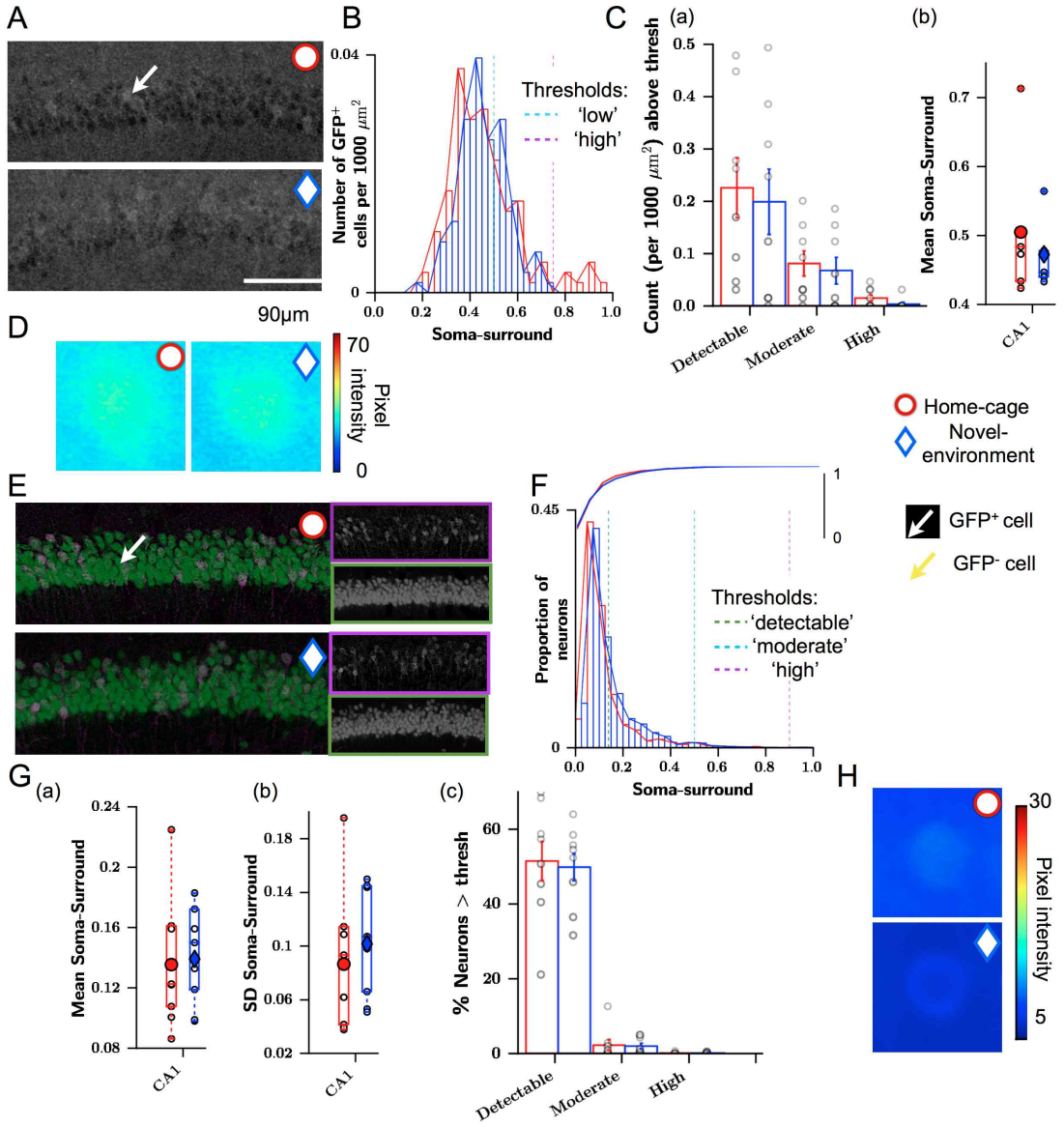


Figure 6.11: *Novel exploration does not significantly alter GFP expression in hippocampal region CA1.* (A) Example confocal 20x max. projection images of CA1 showing the difference in direct GFP fluorescence for a representative pair of home-cage (red circle, upper) and novel-environment (blue diamond, lower) mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 µm^2 . (C) Direct GFP summary measures: (a) The number, in total and exceeding a moderate and high threshold, of GFP$^+$ cells per 1000 µm^2 is summarised across all home-cage (red,*n=8*) and novel-environment (blue, *n=8*) mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all home-cage and novel-environment animals. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron (using the corresponding NeuN labelling), calculated across all detected cells across all home-cage and novel-environment animals. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

I measured GFP labelling in the dentate gyrus using a different method to that for other regions. The dentate gyrus occupies a large region, cellular labelling is clearly distinct from background and cells are sparsely labelled and therefore clearly separable even when imaged in a single plane. As such, I imaged the entire hippocampal region at 4x magnification using widefield illumination and directly quantified cellular expression using the GFP antibody labelling. Widefield images from a home-cage (hc) mouse and a novel-environment (ne) mouse from the same pair are shown in (Fig.6.12(A)). It is immediately evident that neurons in the granule cell layer of the dentate gyrus are strongly labelled with GFP, and therefore express Arc, both at baseline and following exploration. Both the somata and the dendrites of cells are clearly labelled with GFP. The histograms in Fig.6.12(B) show the number of neurons detected per 100 μm of the granule cell layer, assuming that its width remains relatively constant. A large number of neurons can be detected in both conditions, and the distributions are largely overlapping, with a slight bias towards higher soma-surround scores for novel-environment mice. There are, however, no statistically significant differences in the number of neurons detected in total, above a moderate threshold, or above a high threshold (Fig. 6.12 (C)(a): Low: $F_{(1,7)} = 4.89, p = 0.063, Mod : F_{(1,7)} = 5.459, p = 0.05, High : F_{(1,7)} = 1.108, p = 0.33, n = (HC : 4, NE : 5)$) or in the mean soma-surround score for the neurons (Fig. 6.12 (C)(b): $F_{(1,6)} = 0.72, p = 0.43, n = (HC : 4, NE : 5)$).

In summary, exploration of a novel environment does not cause the expected result of increased GFP labelling in the CA1 pyramidal layer. It also has no clear effect on GFP labelling within GFP^+ cells in the dentate gyrus, but it is possible that the number of detectable cells increases following novelty. Even if this is the case, any increase is small relative to the number of neurons that are already expressing Arc under baseline conditions.

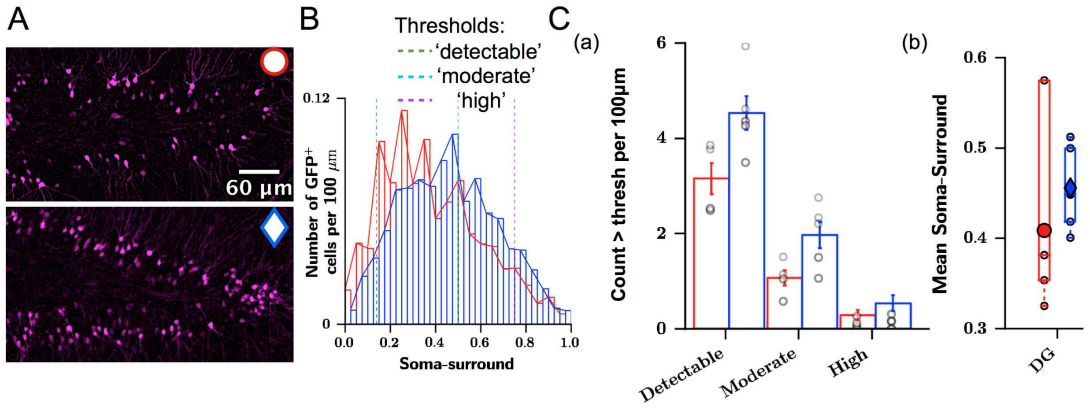


Figure 6.12: *Novel exploration does not significantly increase GFP expression in the hippocampal dentate gyrus region.* (A) Widefield images at 4x magnification showing GFP antibody fluorescence labelling in the dentate gyrus for a home-cage (red circle) and a novel-environment (blue diamond) mouse. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using GFP antibody labelling. Cell counts are given as the density of cells per 100 μm along the granule cell layer. (C) Soma-surround measures for cells detected using GFP antibody labelling in the dentate gyrus: (a) The number of GFP^+ cells per 100 μm in total, exceeding a moderate and a high threshold, (b) Mean soma-surround score

6.4.5 *Exposure to a novel environment increases Arc expression in visual cortex*

6.4.5.1 *GFP expression moderately increases in superficial visual cortex following novelty*

Novel exploration appears to have a small effect on direct GFP fluorescence in superficial visual cortex (Fig. 6.13(A)), shifting the distribution of soma-surround scores of neurons to higher values (Fig. 6.13(B-C)). Across measures, there is a significant effect of condition (MANOVA: $F_{(1,14)} = 5.63, p = 0.010, n = (HC : 8, NE : 8)$). However, due to high variability between animals, novel exploration does not significantly change the proportion of neurons that exceed threshold expression (Fig. 6.13(D)(a): $Total : p = 0.070, Mod : p = 0.185, High : p = 0.182$), but it increases the average intensity of expression within cells (Fig. 6.13(D)(b); p = 0.004).

GFP labelling of NeuN-defined neurons also appears to increase following novel exploration (Fig. 6.13(E-F)), but only in a subset of pairs. Overall, there is no effect of condition (MANOVA: $F_{(1,14)} = 1.06, p = 0.434, n = (HC : 8, NE : 8)$), despite evidence from several measures of a small increase in GFP expression due to novelty (Fig. 6.13(G-H): $Detectable : p = 0.628, Mod : p = 0.182, High : p = 0.309, Mean : p = 0.221$)).

6.4.5.2 *GFP expression moderately increases in visual cortex layer VI following novelty*

In visual cortex layer VI cells appear to be more numerous following exploration of a novel environment (Fig. 6.14(A)). The averaged distributions of somatic GFP labelling across all detected neurons confirm this observation (Fig. 6.14(B)). Detectable GFP^+ cells are more numerous following novel exploration, and the distributions are shifted to the right, suggesting that mean somatic GFP labelling also increases (Fig. 6.14(B)). Across the groups there is a significant increase in the total number of cells detected, and the number of cells exceeding moderate and high thresholds (Fig. 6.14(C)(a): $Count : p = 0.004, Mod : p = 0.005, High : p = 0.027, n = (HC : 8, NE : 8)$. Novel exploration also increases the mean somatic labelling across all cells (Fig. 6.14(C)(b): $p = 0.027, n = (HC : 8, NE : 8)$).

Images showing GFP labelling of NeuN-defined neurons suggest a similar increase in somatic labelling following novel exploration (Fig. 6.14(E)). The distribution of somatic GFP labelling within neurons in visual cortex layer VI shifts to increased intensities following novel exploration (Fig. 6.14(F,H)). However, there are no significant increases across the group for any measure (Fig. 6.14(G), $Detectable : p = 0.734, Mod : p = 0.096, High : p = 0.168, Mean : p = 0.177, SD : p = 0.124, n = (HC : 8, NE : 8)$).

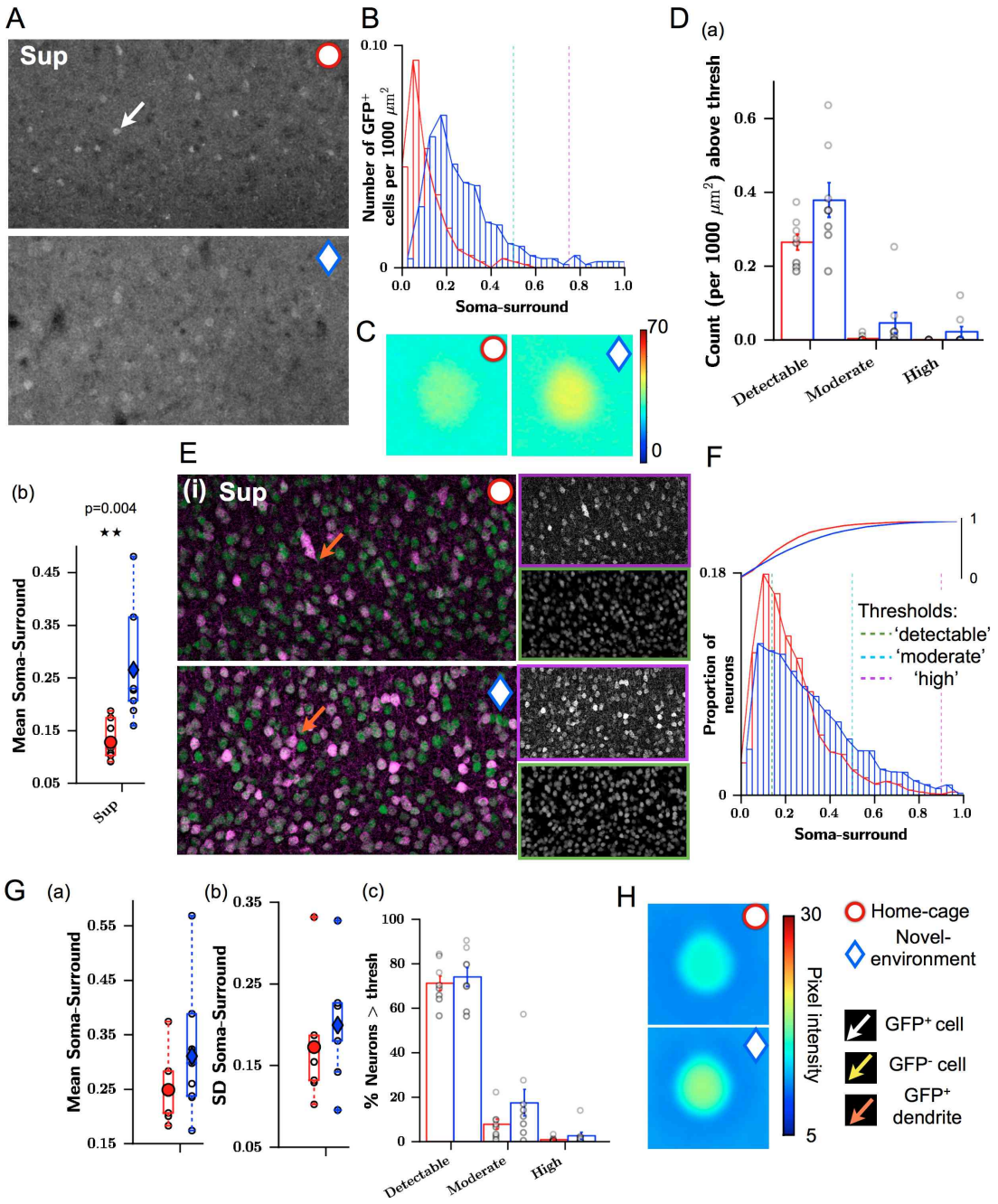


Figure 6.13: *Novelty exposure modestly increases GFP expression in visual cortex* (A) Example confocal 20x maximum projection images of superficial visual cortex showing direct GFP fluorescence for a representative pair of mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 μm^2 . (C) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all home-cage and novel-environment animals. (D) Direct GFP summary measures: (a) The number, in total and exceeding a moderate and high threshold, of GFP^+ cells per 1000 μm^2 .(b) The mean soma-surround of detected cells. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all neurons detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected (NeuN) neuron. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

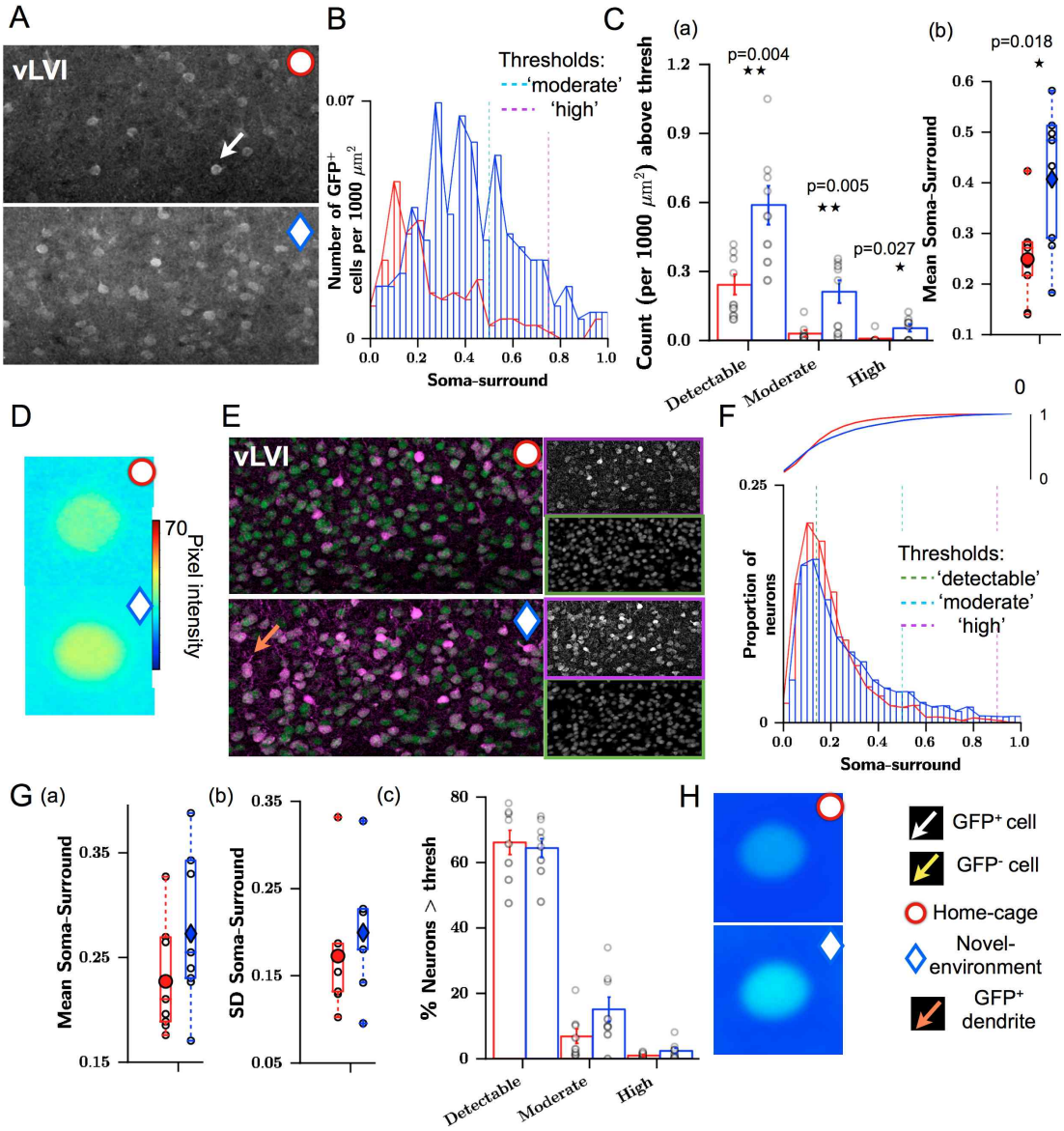


Figure 6.14: *Novelty exposure increases GFP expression in visual cortex layer VIa* (A) Example confocal 20x maximum projection images of visual cortex layer VI showing the difference in direct GFP fluorescence for a representative pair of mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 μm^2 . (C) Direct GFP summary measures: (a) The number, in total and exceeding a moderate and high threshold, of GFP^+ cells per 1000 μm^2 is summarised across all home-cage (red,*n=8*) and novel-environment (blue, *n=8*) mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected neuron, calculated across all detected cells across all mice. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron (using the corresponding NeuN labelling), calculated across all detected cells across all mice. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.4.5.3 *Arc protein expression in visual cortex layer VI increases following novelty*

Data from direct GFP fluorescence measures suggest that Arc-GFP activation increases following novelty exposure in layer VI of visual cortex. Here, I have investigated whether protein expression follows Arc-GFP activation. Images from a home-cage and novel-environment animal show that novel exploration considerably increases the number of detectable Arc-labelled cells in this example, as well as increasing the number of detectable dendrites (Fig. 6.15 (A)). The expression of Arc in the average detected cell also increases (Fig. 6.15 (B)).

Across animals, novelty increases the number of cells expressing low, moderate and high levels of *Arc* protein (Fig. 6.15 (C)(CD)(a): $\mathrm{Low} : p = 0.007, \mathrm{Mod} : p = 0.030; \mathrm{High} : p = 0.035, n = (3, 6)$) and the level of *Arc* protein antibody labelling within cell somata relative to their surround (Fig. 6.15 (B),(C)(a): $p = 0.015, n = (3, 6)$).

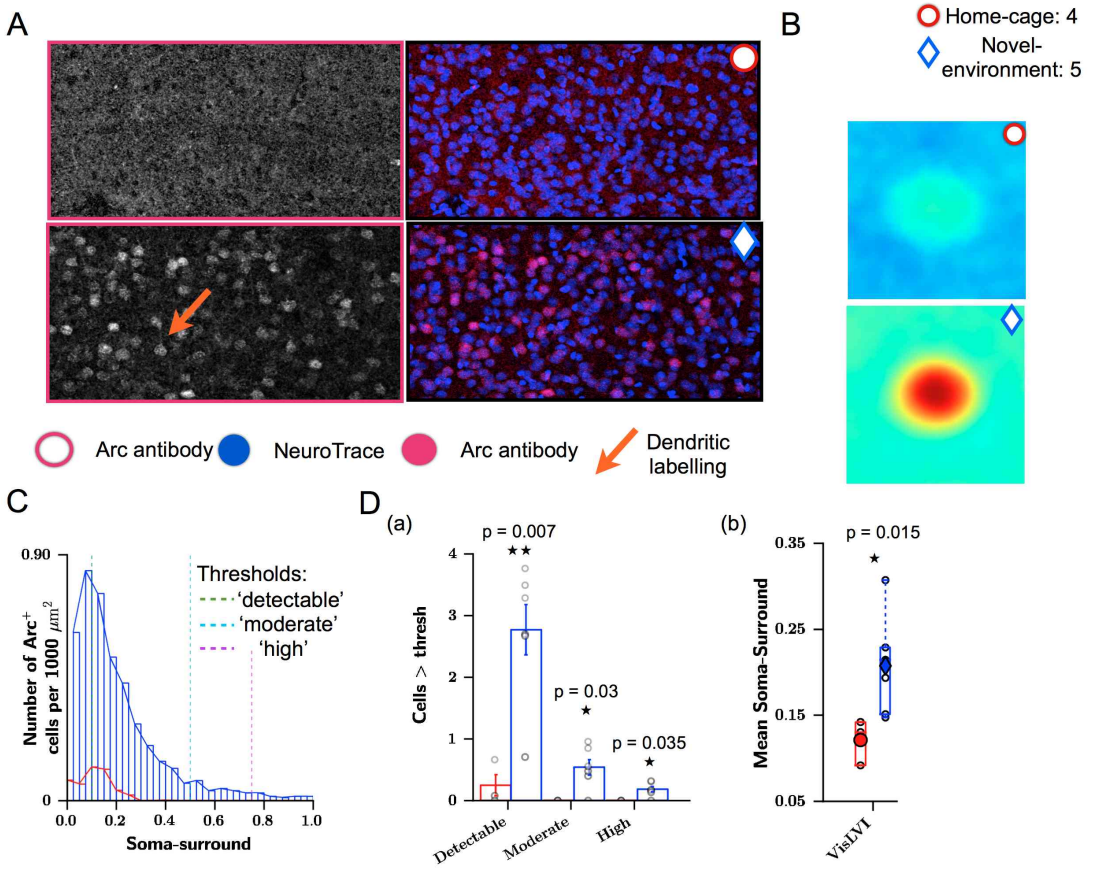


Figure 6.15: *Novelty exposure increases Arc protein labelling in visual cortex* (A) Example confocal 20x maximum projection images showing Arc protein labelling (magenta outline) for a representative pair of home-cage (red circle, upper) and novel-environment (blue diamond, lower) mice. Arc protein was co-labelled with NeuroTrace (blue). (B) Histograms show the distributions of Arc soma-surround labelling fluorescence scores for all detected cells. (C) Soma-surround measures for Arc antibody labelling detected cells: (a) the number of cells with a soma-surround score exceeding a low, moderate, or 'high' threshold, (b) Mean soma-surround score. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.05$, *** $p < 0.01$.

6.4.6 *Behavioural modulation of Arc -GFP activation in the absence of Arc protein*

In the previous sections I have shown that in Arc^+/GFP^+ animals, exposure to a novel environment up-regulates both *Arc* -GFP activation and *Arc* translation, particularly in MEC deep layers and, to a lesser degree, in the visual cortex. Here, I have considered whether GFP^+/GFP^+ mice, which do not express *Arc* protein, show similar up-regulation in *Arc* -GFP activation.

6.4.6.1 *Novel exploration increases Arc -GFP activation in MEC layer II in GFP^+/GFP^+ mice*

Direct fluorescence measures: In GFP^+/GFP^+ mice, it appears to be easier to detect GFP^+ cells (Fig. 6.16(A)) than in Arc^+/GFP^+ mice, which only possess one copy of GFP DNA. Across both ventral and dorsal regions, the density of cells detected remains relatively low but, in contrast to data shown for Arc^+/GFP^+ mice, somatic expression appears to be slightly higher following novel exploration than at baseline (Fig. 6.16(B)). Indeed, novel exploration increases the number of cells exceeding a moderate threshold (Fig. 6.16(C)(b): ANCOVA: $F_{(1,14)} = 4.14, p = 0.047, n = (HC : 4, NE : 5)$), and the mean soma-surround of all cells (Fig. 6.16(C)(c): ANCOVA: $F_{(1,14)} = 4.30, p = 0.044, n = (HC : 4, NE : 5)$), but not the overall number of cells detected (Fig. 6.16(C)(a), $F_{(1,14)} = 2.10, p = 0.15$).

These measures are not constant along the DV axis of MEC layer II. I demonstrated that in Arc^+/GFP^+ mice, there is a very shallow gradient in expression from dorsal to ventral regions that is not significantly modulated by novel exploration. In GFP^+/GFP^+ mice, there is also a significant effect of region across animals in both conditions for cell count in total ($F_{(6,55)} = 3.85, p = 0.0027$), and exceeding a moderate threshold ($F_{(6,55)} = 7.87, p < 0.0001$), and the mean soma-surround measure ($F_{(6,46)} = 3.05, p = 0.014$)(Fig. 6.16(C)). Linear regression analysis reveals that the slopes of the gradients are shallow (Fig. 6.16(D), Count: slope = -0.27, p = 0.0178). There is no interaction between group and region on any measure (ANCOVA: $p < 0.7$).

GFP antibody labelling measures: Novel exploration also increases almost all measures of GFP antibody labelling, which is particularly evident in dorsal regions (Fig. 6.17 (A-B)). There is an increase in mean soma-surround (Fig. 6.17 (C)(a): $F_{(1,56)} = 22.4, p < 0.0001, n = (HC : 4, NE : 5)$), variability in soma-surround (Fig. 6.17 (C)(b): $F_{(1,56)} = 17.2, p = 0.00010, n = (HC : 4, NE : 5)$), and the proportion of neurons exceeding a moderate or high threshold (Fig. 6.17 (C)(d-e): Mod: $F_{(1,56)} = 24.4, p < 0.00001$,High: $F_{(1,56)} = 9.43, p = 0.0032, n = (HC : 4, NE : 5)$). There is no effect on the number exceeding a low threshold ($F_{(1,56)} = 3.75, p = 0.027$).

Again, I tested for the presence of a gradient. Across both groups, from the dorsal to ventral extent the variability in somatic GFP labelling decreases (Fig. 6.17(C)(b): $F_{(7,56)} = 5.30, p < 0.00001$) and the proportion of neurons exceeding a moderate threshold decreases (Fig. 6.17(C)(d): $F_{(7,56)} = 3.72, p = 0.0019$). The mean somatic GFP labelling shows a non-significant downward trend along the axis (Fig. 6.17(C)(a): $F_{(7,56)} = 2.51, p = 0.099$). There is no interaction between group and region on any measure, although there is a close to significant interaction for the mean somatic expression (ANCOVA: $F_{(1,7,56)} = 1.87, p = 0.0915$)

To test whether there is a systematic effect of dorso-ventral distance on GFP expression in novel-environment mice, I performed a linear regression analysis. Mean somatic expression and the variability in somatic expression decrease significantly (Fig. 6.17(D)(a-b): Mean $slope = -2.7(pixelintensity/mm), r^2 = 0.42, p < 0.0001$, SD $m = -2.6, r^2 = 0.48, p < 0.0001$). The proportion of neurons with moderate GFP labelling also decreases significantly (Fig. 6.17(D)(c): $m = -0.15, r^2 = 0.21, p = 0.0035$). This means that the proportion of moderately GFP expressing neurons decreases by ~ 15% over 1mm of the DV extent.

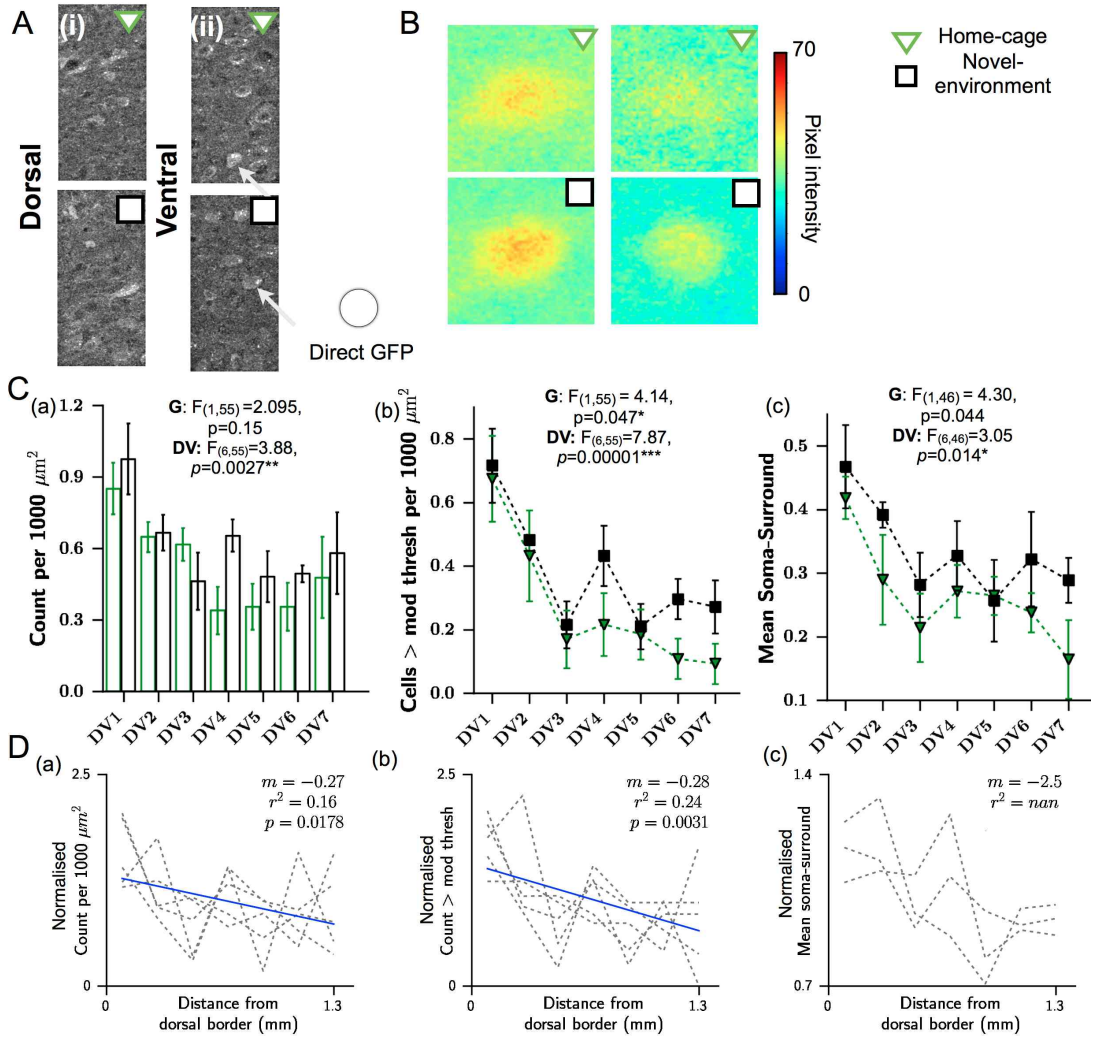


Figure 6.16: *A DV gradient in somatic GFP fluorescence in GFP$^+$/GFP$^+$ animals that express no functional Arc protein.* (A) Example confocal images of direct GFP fluorescence in dorsal and ventral MEC layer II for a representative pair of mice from each group. (B) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all home-cage and novel-environment animals. (C) Summary measures: The number of cells (a) in total, (b) exceeding a moderate threshold is shown per 1000 μm ^{2} region. (c) Mean soma-surround score for detected cells. (D) Linear regression plots show the features in (C) for novel-environment animals plotted as a function of DV extent. Blue line indicates the line of best fit, where *m* is the gradient, and r^2 is calculated using data normalised by the individual mean for each animal. Grey lines indicate individual results. The results of a 2-way ANCOVA are indicated for condition (G) and region (DV).

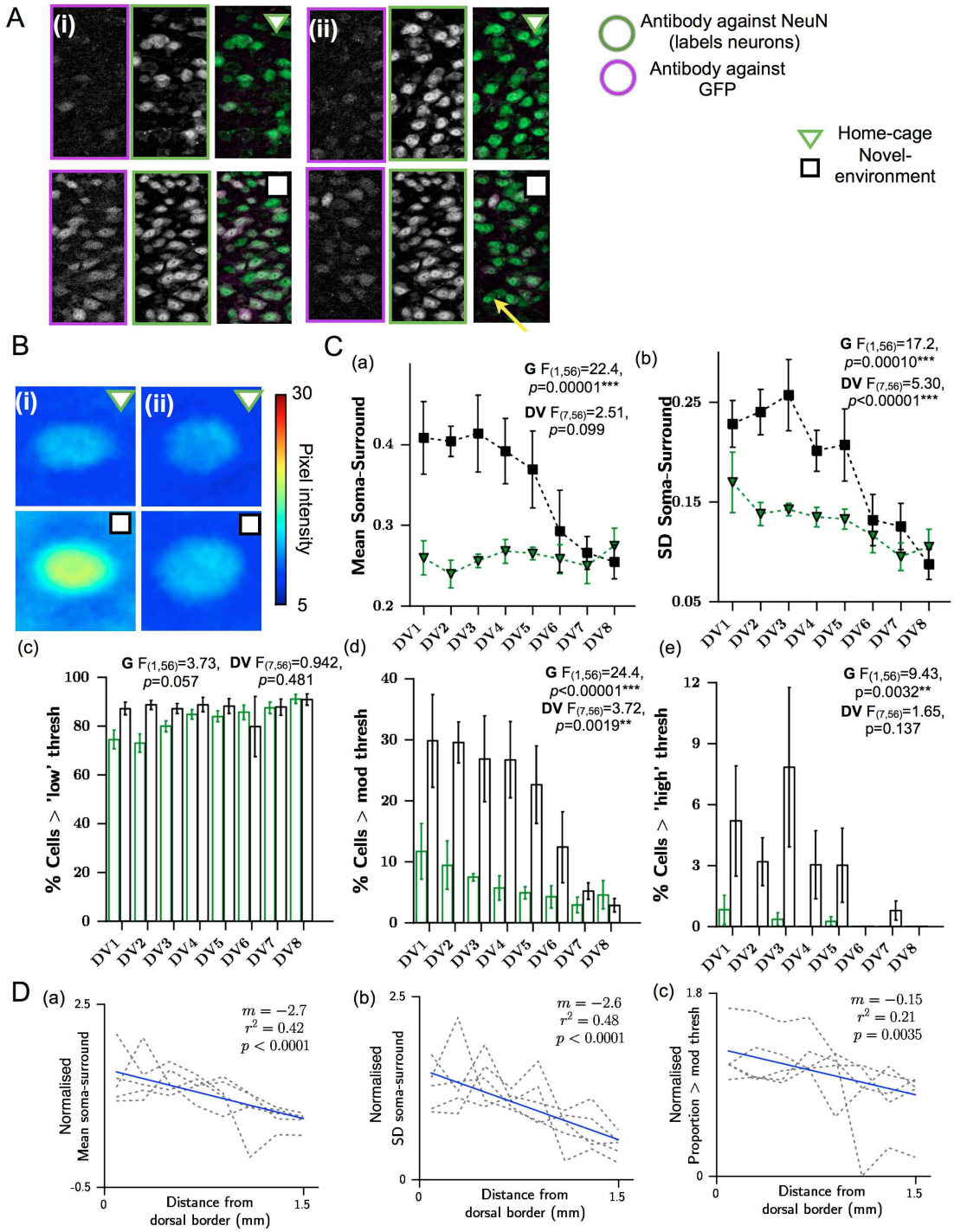


Figure 6.17: *A DV gradient in somatic GFP antibody expression in GFP$^+$/GFP$^+$ animals that express no functional Arc protein.* (A) Example confocal images of GFP antibody labelling and NeuN in dorsal and ventral MEC layer II for a representative pair of mice from each group. (B) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a NeuN-detected neuron, calculated across all detected cells across all home-cage and novel-environment animals. (C) Summary measures for all detected neurons: (a) Mean soma-surround score, (b) standard deviation in soma-surround score, (c-e) The proportion of neurons exceeding a (a) low, (b) moderate, (c) high threshold. (D) Linear regression plots show the features in (C) for novel-environment animals. Blue line indicates the line of best fit, where m is the gradient (for data that has not been normalised), and r^2 is calculated using data normalised by the individual mean for each animal. Grey lines indicate individual results. ANCOVA results are indicated for group (G) and region (DV). HC: n = 4, NE: n = 5.

6.4.6.2 *Novel exploration continues to increase Arc -GFP activation in MEC deep layers in GFP^+/GFP^+ mice*

In MEC deep layers, there is a clear increase in GFP expression following novel exploration that closely parallels the data shown for Arc^+/GFP^+ animals (Fig. 6.8). Example images show a noticeable increase in the number and intensity of directly detectable GFP^+ cells following novel exploration (Figs. 6.18(A-B)). Consistent with the images shown, across the group, novel exploration increases the number of directly detectable cells (Fig. 6.18 (C)(a): p=0.025, n=(HC:4,NE:5)), their mean soma-surround score (Fig. 6.18 (C)(b): p=0.026, n=(HC:4,NE:5)) and the number of neurons exceeding a moderate threshold (Fig. 6.18 (C)(a): p=0.0061, n=(HC:4,NE:5)). This change can be visualised in Fig. 6.18(D).

The results for GFP antibody labelling are highly similar and emphasise how robust the increase in Arc-GFP activation in MEC deep layers is. Novelty increases the visual prevalence of GFP^+ neurons (Fig. 6.18 (E)), it changes the shape of the distribution of soma-surround scores across all neurons(Fig. 6.18 (F)), and it follows that it increases the mean neuronal soma-surround GFP labelling (Fig. 6.18 (H)(a): p=0.012, n=(HC:4,NE:5)), the variability between cells (Fig. 6.18 (H)(b): p=0.0097, n=(HC:4,NE:5)) and the proportion of cells exceeding a low and moderate threshold (Fig. 6.18 (H)(c): p=0.046, p = 0.023, n=(HC:4,NE:5)). The group data suggest that the proportion of moderately expressing neurons increases from less than 5% to 20% (Fig. 6.18 (F)(c)). The difference between the average cell at baseline and following exploration can be seen in Fig. 6.18 (G).

In summary, Arc protein is not required for the robust up-regulation of Arc-GFP activation evident in MEC deep layers cells following exploration of a novel environment.

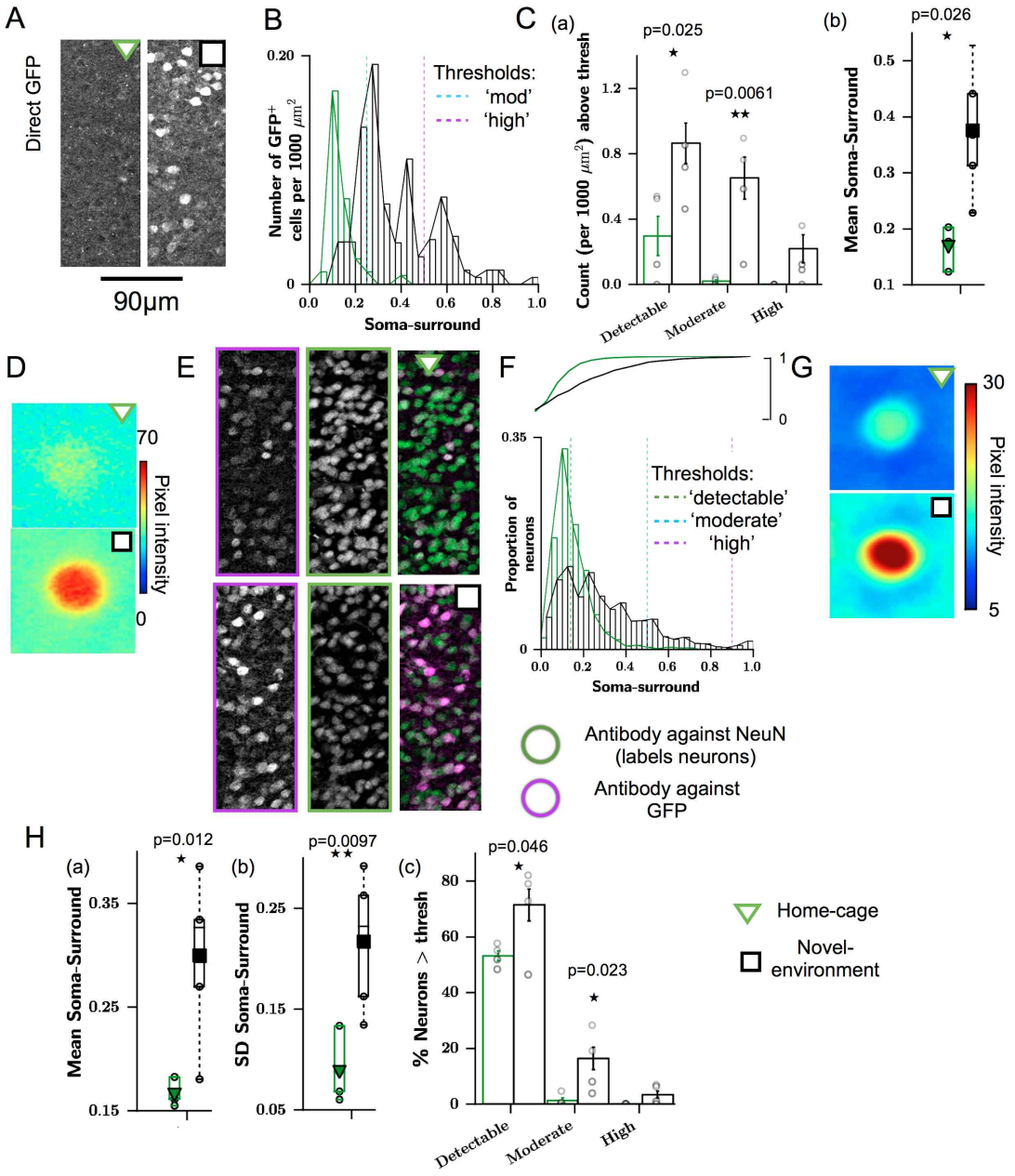


Figure 6.18: *Novel exploration increases GFP expression in MEC deep layers in Arc KO animals.* (A) Example confocal 20x maximum projection images of dorsal MEC deep layers showing direct GFP fluorescence for a representative pair of mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 µm ^{2} . (C) Direct GFP measures: (a) The number of GFP$^+$ cells per 1000 µm ^{2} , and number exceeding a moderate and high threshold are summarised across all home-cage (green,*n=4*) and novel-environment (black, *n=5*) mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice in each group. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron, calculated across all detected cells across all mice in each group. (H) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a low, moderate and 'high' threshold (see F). Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.4.6.3 *Novel exploration does not significantly increase Arc -GFP activation in the putative parasubicular patches in GFP$^+$/GFP$^+$ mice*

The large patches dorsal to MEC layer II do not show detectable increases in Arc-GFP activation following novel exploration in Arc$^+$/GFP$^+$ mice. This is also the case in GFP$^+$/GFP$^+$ mice, where novel exploration has no significant effect on any direct GFP fluorescence or GFP antibody labelling measure. Images of direct fluorescence show that both in baseline and novel-environment animals, neurons expressing high levels of GFP can be detected (Fig.6.19(A)). The distribution of soma-surround scores for the two groups are largely overlapping (Fig.6.19(B)) and neither the number of neurons detected nor the mean soma-surround varies consistently (Fig.6.19(C), $\mathtt{Count} : p = 0.283, \mathtt{Mod} : p = 0.645, \mathtt{High} : p = 0.759, \mathtt{Mean} : p = 0.386$). Detected cells in the two groups appear highly similar (Fig.6.19(D)).

Measurements with GFP antibody labelling appear to show more evidence for an effect of novelty; a larger proportion of neurons are labelled (Fig.6.19(E)), the distribution of soma-surrounds shifts (Fig.6.19(F)), and for all measurements taken for soma-surround score (Fig.6.19(G)), novelty appears to show higher values ($\mathtt{Detectable} : p = 0.137, \mathtt{Mod} : p = 0.058, \mathtt{High} : p = 0.190, \mathtt{Mean} : p = 0.085, \mathtt{SD} : p = 0.114$). This can be visualised in Fig. 6.19(H)). In fact, the pattern closely resembles results for dorsal MEC layer II, where Arc$^+$/GFP$^+$ mice and GFP$^+$/GFP$^+$ mice at baseline appear to show similar levels of Arc-GFP activation whereas novel exploration in the latter increases Arc-GFP activation. However, the variability between animals within the same group of GFP$^+$/GFP$^+$ mice, combined with the low number, appears to override that between groups, so there is insufficient evidence to conclude that in GFP$^+$/GFP$^+$ mice, we see a different pattern of results.

6.4.6.4 *Novel exploration has no effect on Arc -GFP activation in CA1 in GFP$^+$/GFP$^+$ mice*

One of the more surprising results presented for Arc$^+$/GFP$^+$ mice was that CA1 cells do not increase Arc-GFP activation during novel exploration. For direct fluorescence measures in GFP$^+$/GFP$^+$ mice I find very similar GFP levels between the baseline group and exploration group, which adds evidence to this finding. Example images look very similar (Fig. 6.20(A)) and the distributions of soma-surround score are highly overlapping (Fig. 6.20(B)). As such, there are no differences in the number of cells detected or in the mean soma-surround GFP across detected cells (Fig. 6.20(C): $\mathtt{Count} : p = 0.878, \mathtt{Mod} : p = 0.869, \mathtt{High} : p = 0.767, \mathtt{Mean} : p = 0.948$). There is a similar absence of change in Arc-GFP activation using the GFP antibody labelling as a marker (Fig. 6.20(D-H), $\mathtt{Detectable} : p = 0.921, \mathtt{Mod} : p = 0.237, \mathtt{High} : p = 0.052, \mathtt{Mean} : p = 0.274$), except for neuronal variability ($\mathtt{SD} : p = 0.041$).

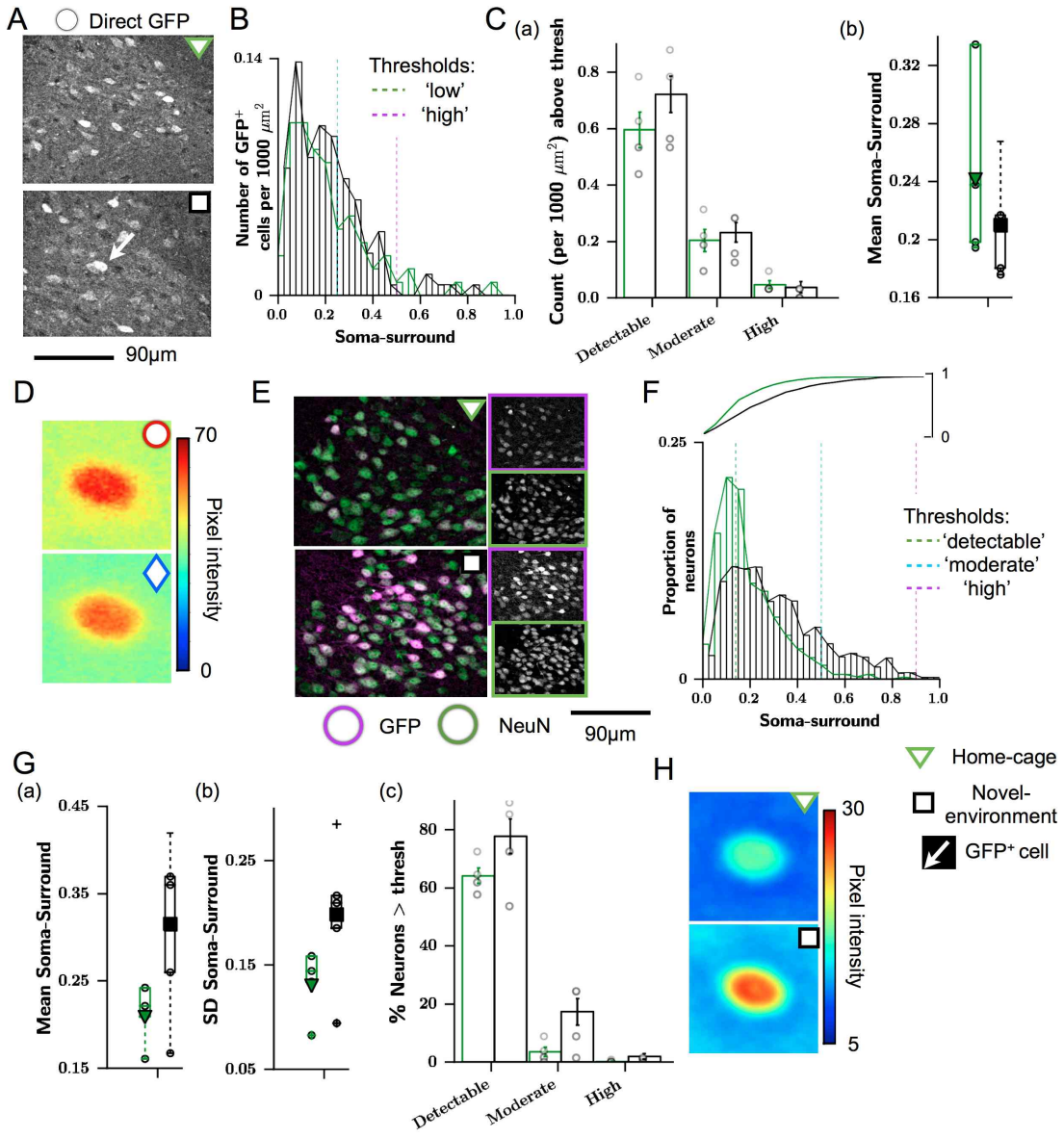


Figure 6.19: *GFP labelling in the large patches dorsal to MEC* (A) Example confocal 20x maximum projection images of the region of cell patches dorsal to MEC layer II showing direct GFP fluorescence for a representative pair of home-cage and novel-environment mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 µm^2 . (C) Direct GFP measures: (a) The number of GFP$^+$ cells per 1000 µm^2 , and number exceeding a moderate and high threshold are summarised across all home-cage (*n*=4) and novel-environment *n*=5) mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice in each group. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a low, moderate and 'high' threshold (see F). (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron (using the corresponding NeuN labelling), calculated across all detected cells across all mice in each group. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

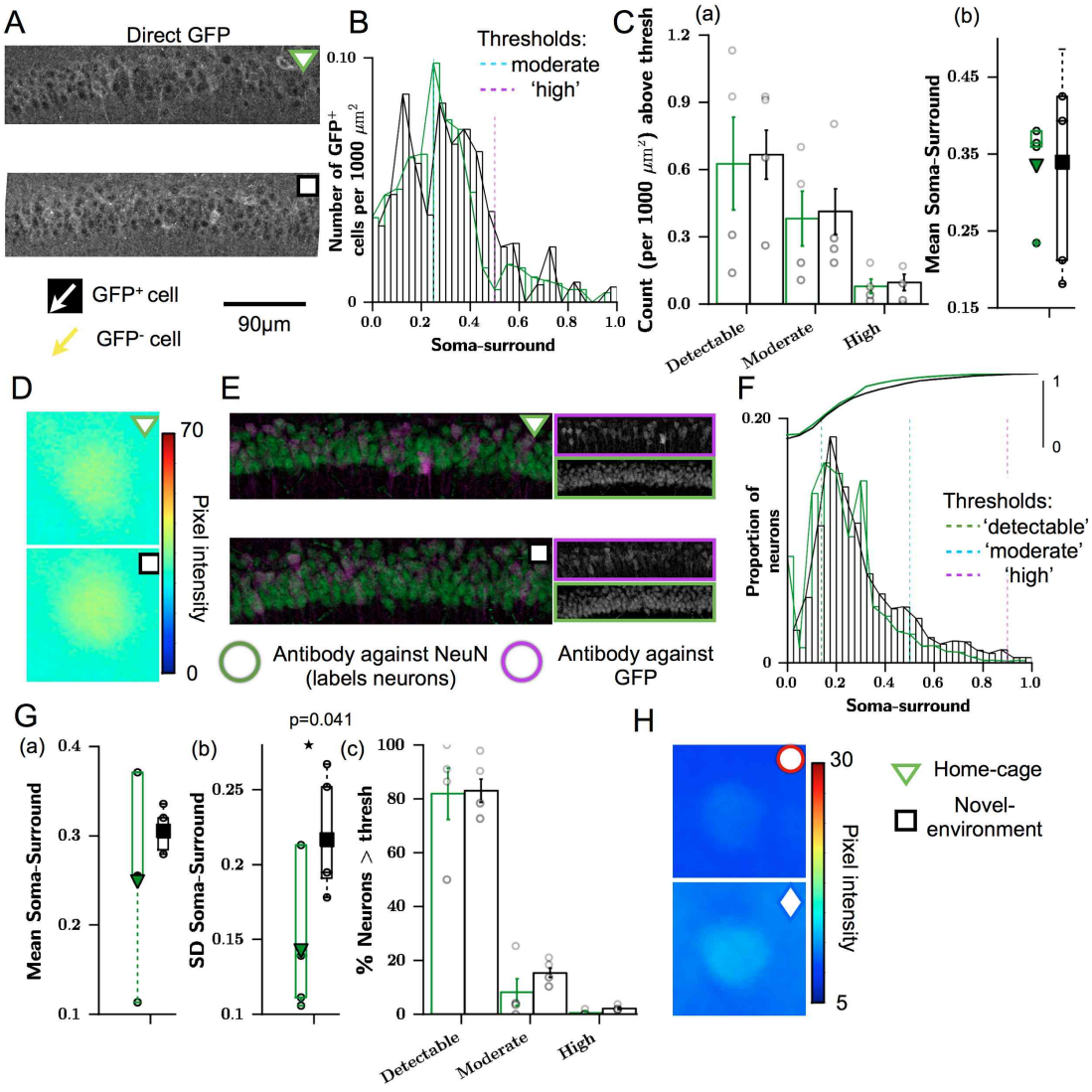


Figure 6.20: *Novel exploration does not significantly alter GFP expression in hippocampal region of CA1 in Arc KO mice.* (A) Example confocal 20x maximum projection images CA1 showing the difference in direct GFP fluorescence for a representative pair of home-cage and novel-environment mice. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 µm ^{2} . (C) Direct GFP measures: (a) The number of GFP$^+$ cells per 1000 µm^2 , and number exceeding a moderate and high threshold are summarised across all home-cage (green,*n=8*) and novel-environment (black, *n=8*) mice. (b) The mean soma-surround of detected cells. (D) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice in each group. (E) As (A), but showing GFP antibody labelling (magenta outline), NeuN antibody labelling (green outline) and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a low, moderate and 'high' threshold (see F). (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron calculated across all detected cells across all mice in each group. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.4.6.5 *Novel exploration does not consistently increase Arc -GFP activation in visual cortex superficial layers in GFP$^+$/GFP$^+$ mice*

In Arc$^+$/GFP$^+$ mice, novel exploration modestly increases GFP expression in visual cortex, when measured using direct fluorescence. There is no significant effect of novel exploration in visual cortex of GFP$^+$/GFP$^+$ mice, despite evidence from example pairs (Fig. 6.21(A)) and a shift in the distribution (Fig. 6.21(B-D)). However, the variability across animals is considerable (Fig. 6.21(C): $\texttt{Count} : p = 0.698, \texttt{Mod} : p = 0.490, \texttt{High} : p = 0.526, \texttt{Mean} : p = 0.601$).

The results for GFP antibody labelling closely mirror direct fluorescence results (Fig. 6.21(E-G): $\texttt{Detectable} : p = 0.835, \texttt{Mod} : p = 0.092, \texttt{High} : p = 0.121, \texttt{Mean} : p = 0.051, \texttt{SD} : p = 0.090$). One noticeable effect is that novel exploration appears to increase the labelling of neuropil surrounding cells in superficial cortex (Fig. 6.21(H)).

6.4.6.6 *Novel exploration consistently increases Arc -GFP activation in visual cortex deep layers in GFP$^+$/GFP$^+$ mice*

In visual cortex layer VI, the effect of novel exploration can be robustly detected. There is an increase in the number of visually detectable neurons using direct GFP fluorescence (Fig. 6.22 (A)), accompanied by a shift in the soma-surround score of detectable neurons to higher intensities (Fig. 6.22 (B)). The mean soma-surround score increases significantly (Fig. 6.22 (D)(a): p=0.0063, n=(HC:4,NE:5)). The total number of detected cells, and number exceeding moderate and high thresholds also increase (Fig. 6.22 (D)(b): p=0.0068, p = 0.003, p = 0.04, n=(HC:4,NE:5)). A similar pattern is evident using GFP antibody labelling, where the distribution of neurons becomes considerably wider (Fig. 6.22 (E-F)). The mean soma-surround and variability in soma-surround of neurons increases (Fig. 6.22 (D)(a): p=0.0042, p = 0.0033, n=(HC:4,NE:5)). Notably, novel exploration increases the number of moderately expressing neurons from none to 20% (Fig. 6.22 (D)(b): Mod: p=0.011, High: p = 0.070, n=(HC:4,NE:5)).

In summary, in GFP$^+$/GFP$^+$ mice, novel exploration increases Arc-GFP activation in visual cortex without the requirement for Arc protein. This effect is only robust in layer VI, and is highly consistent but perhaps more robust than results from Arc$^+$/GFP$^+$ mice.

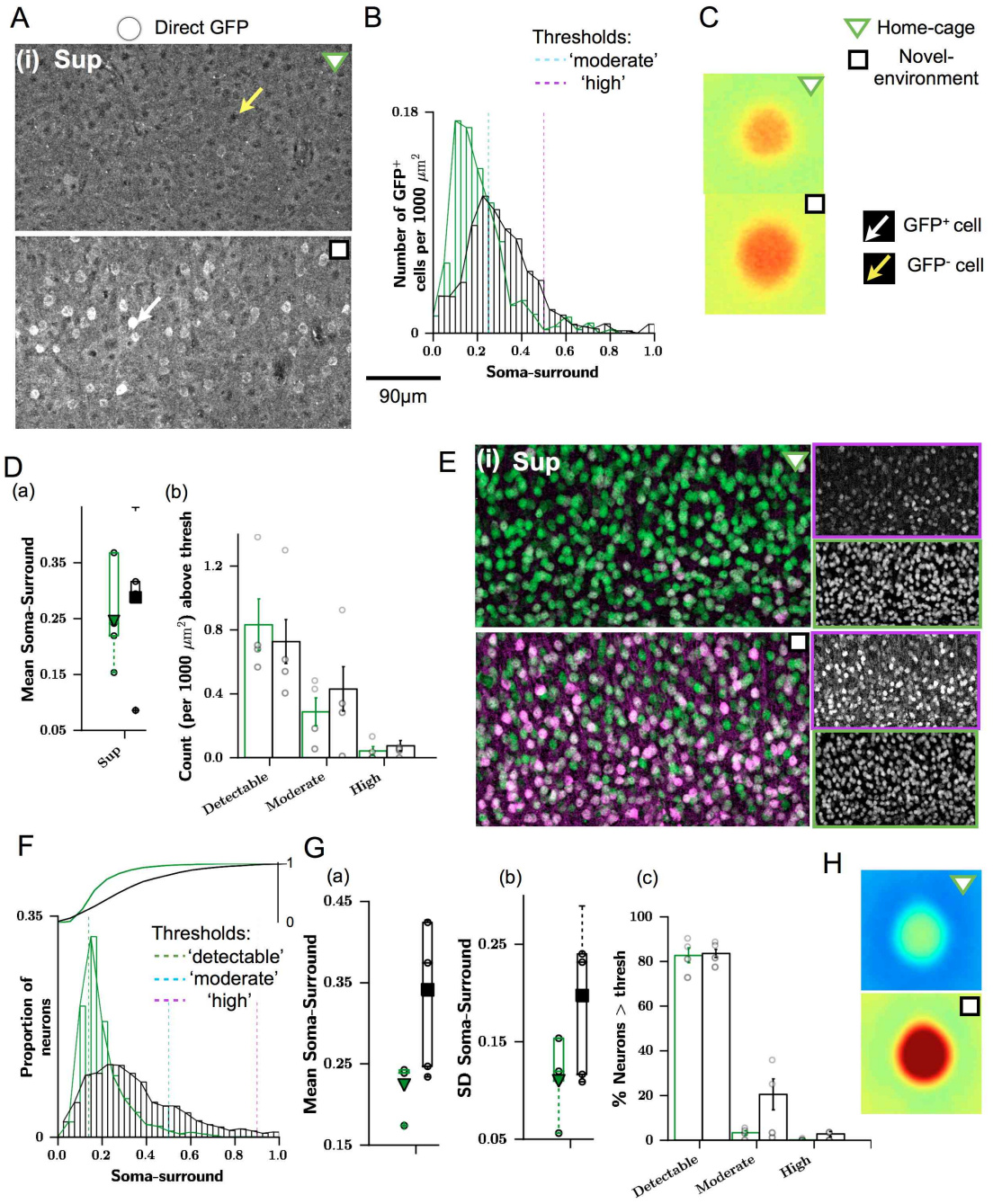


Figure 6.21: *The effect of novel exploration on directly imaged GFP fluorescence in superficial visual cortex* (A) Example confocal 20x max. projection images of visual cortex superficial layers showing direct GFP fluorescence for a representative pair. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 μm^2 . (C) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice. (D) Direct GFP summary measures: (a) The mean soma-surround of detected cells. (b) The number, in total and exceeding a moderate and high threshold, of GFP^+ cells per 1000 μm^2 is summarised across all mice. (E) As (A), but showing GFP antibody labelling, NeuN antibody labelling and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron, calculated across all detected cells across all mice. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

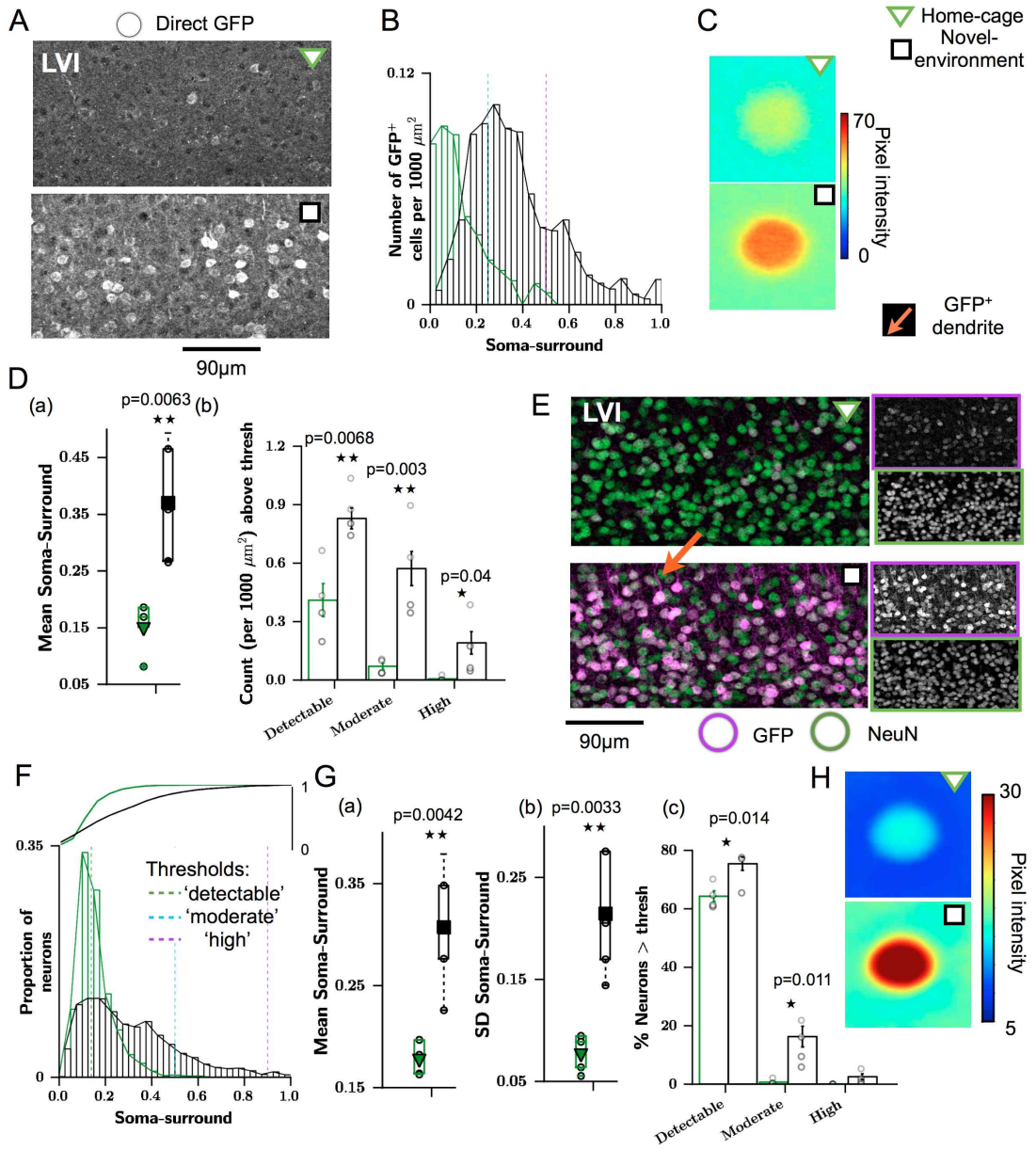


Figure 6.22: *Novelty exposure increases GFP expression in visual cortex deep layers in animals that express no functional Arc protein.* (A) Example confocal 20x maximum projection images of visual cortex layer VI showing direct GFP fluorescence for a representative pair. (B) Histograms show the averaged distributions of soma-surround scores for all cells detected using direct GFP fluorescence. Cell counts are given as the density of cells per 1000 μm^2 . (C) Images show the average 60x60 pixel region of direct GFP fluorescence surrounding a detected cell, calculated across all detected cells across all mice. (D) Direct GFP summary measures: (a) The mean soma-surround of detected cells. (b) The number, in total and exceeding a moderate and high threshold, of GFP^+ cells per 1000 μm^2 is summarised across all mice. (E) As (A), but showing GFP antibody labelling, NeuN antibody labelling and GFP antibody (magenta)/NeuN (green) overlay. (F) Histograms show the averaged distributions (lower) and cumulative distributions (upper) of soma-surround scores for GFP antibody labelling across all cells detected using NeuN. (G) Soma-surround measures for GFP antibody labelling within NeuN-detected neurons: (a) Mean, (b) standard deviation, (c) the percentage of neurons with a soma-surround score exceeding a 'low','moderate', and 'high' threshold. (H) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a detected neuron, calculated across all detected cells across all mice. Asterisks indicate one-way ANOVA significance alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

6.5 DISCUSSION

In this chapter I have investigated how novel exploration influences *Arc* expression in medial entorhinal cortex (MEC), hippocampus and visual cortex. By quantifying the direct fluorescence and neuronal cell body expression of GFP, and the expression of Arc protein, in an Arc-GFP mouse line that expresses GFP under the control of the endogenous Arc promoter, I have been able to show that both Arc-GFP activation and Arc protein expression show selective novel-exploration-dependent changes. *Arc* is an activity-dependent gene (Lyford et al., 1995) that is critical for certain forms of plasticity (Plath et al., 2006). Here, I outline how novel exploration has differential effects on *Arc* expression within spatial navigation circuits, then I discuss the implications for neural circuit function.

6.5.1 *The effect of novel exploration on Arc -GFP activation and protein expression throughout spatial navigation circuits*

The most striking and robust effect of novel exploration on *Arc* -GFP activation is observed in the deep layers V/VI of the MEC (Fig. 6.8), which have been shown to provide a major output of the MEC to the neocortex (Insausti et al., 1997). The effect of Arc-GFP activation is mirrored by a large increase in the number of cells expressing detectable Arc protein (Fig. 6.9). It suggests that layer V/VI cells are both active and potentially requiring plasticity-related proteins during novel spatial exploration.

This considerable effect contrasts with a complete absence of detectable change in *Arc* -GFP activation in MEC layer II (Figs. 6.3, 6.5), a layer known to possess cells that are active during novel spatial exploration (Hafting et al., 2005; Burgalossi et al., 2011). Rather, consistent with the data presented in the previous chapter for home-cage animals, somatic expression of Arc in MEC layer II is low and continues to follow a dorso-ventral gradient after novel exploration. This presents a challenge for our understanding of Arc, which is considered as a reporter of neural activity (Lyford et al., 1995). However, Arc protein expression in MEC layer II does increase following novel exploration, suggesting that there may be a dissociation between Arc-GFP activation and protein levels.

The lack of effect of novel exploration on *Arc* -GFP activation in MEC layer II is surprising, but a second surprising result is that *Arc* -GFP activation does not show a detectable increase in the CA1 region of hippocampus (Fig. 6.11), where its levels are very low compared with other regions studied. In addition, *Arc* -GFP activation does not increase significantly in the dentate gyrus, where even at baseline, a significant number of cells express high levels of Arc (Fig. 6.12). These results contrast with previous work showing increases in Arc RNA in the hippocampus (Guzowski et al., 1999). As I will discuss in more detail below, CA1 neurons are spatially modulated and a subset are known to be active when an animal explores its environment (O'Keefe, 1976). It follows that a gene that simply reports activity should show increased expression after exploration, which suggests either that *Arc* requires stricter criteria for activation or that GFP expression can fail to indicate promoter activity.

In addition to exploring changes in *Arc* expression in MEC and the hippocampus, I also measured changes in visual cortex. Previous work, including in the same mouse line, has

shown that Arc-GFP activation increases in visual cortex following exposure to novelty (Ons et al., 2004; Wang et al., 2006). The visual cortex exchanges connections with MEC (Insausti et al., 1997). I found that Arc-GFP activation also increases in layer VI, the deep layer of visual cortex, where it is paralleled by an increase in Arc protein. It also shows tendencies towards increasing in the superficial visual cortex.

The final result to highlight is that *Arc* protein does not appear to be necessary for the up-regulation of *Arc* -GFP activation following exposure to a novel environment as in both MEC and visual cortex deep layers, GFP^+/GFP^+ mice, which possess no *Arc* protein, show the same patterns of changes in *Arc* -GFP activation as Arc^+/GFP^+ mice. This is consistent with the role of Arc as a cytoskeletal protein that does not directly impact on regulation of transcription.

So, what can this pattern of changes in *Arc* -GFP activation and protein expression tell us about Arc and about its role in neural circuits such as the MEC? In the proceeding sections, I discuss the results presented here in the context of the neural circuits under investigation.

6.5.2 *Exploration of a novel environment does not increase Arc -GFP activation in MEC layer II*

In the data presented here there is no evidence that exploration of a novel environment increases Arc-GFP activation in MEC layer II. In fact, Arc expression is strikingly similar following exposure to a novel environment and at baseline throughout the dorso-ventral extent of MEC (Fig. 6.6). Since Arc is typically thought to be activity-regulated (Lyford et al., 1995) and these cells are known to be active during spatial exploration (Hafting et al., 2005; Fyhn et al., 2008; Burgalossi et al., 2011), it is difficult to explain this finding.

The low level and lack of novelty-induced increase in Arc-GFP activation in MEC layer II suggests either that these cells do not regulate Arc expression above low levels, or that the behavioural setting in which the mice were placed is not sufficient to drive Arc-GFP activation. Two pieces of evidence suggest that the latter may be the case. The first evidence is that cells expressing moderate levels of Arc do exist in layer II, even if they cannot be observed across all animals (Fig. 6.3). It is possible that only a subset of neuron types express Arc, particularly as there are multiple cell populations, including stellate cells, pyramidal cells, and fast-spiking interneurons (Jones and Buhl, 1993; Gatome et al., 2010), but expression is extremely sparse.

The second piece of evidence that the lack of Arc-GFP activation may be tied to task requirement rather than the potential of the cells is that the level of Arc protein increases in MEC layer II (Fig. 6.7), indicating that Arc protein expression is correlated with function. Research in rats has shown that Arc protein expression may be particularly sensitive to how novel a task is, as it does not increase at all during lever-pressing in over-trained rats, whereas recorded levels double in newly-trained rats (Kelly and Deadwyler, 2003). Research using high-sensitivity fluorescent ISH has also indicated that Arc transcription, and therefore promoter activity, is not directly coupled to activity in that CA1 neurons show stable firing over repeated presentations of the same novel environments, whereas Arc nuclear RNA is reduced (Guzowski et al., 2006). A more demanding task than exploration of a novel arena, perhaps requiring spatial memory or navigation, would perhaps be necessary to increase activity. Other research has provided limited evidence that Arc RNA expression does not increase in ven-

tral MEC layer II in rats that have either performed a water maze task or swam as a control (Gusev et al., 2005), which again suggests that increases in Arc transcription may have more specific requirements.

I found that there is a shallow gradient in Arc-GFP activation that can be detected both in cage-control animals (see previous chapter on baseline expression) and those that have explored a novel environment (Fig. 6.6). Neurons in dorsal regions show, on average, higher somatic expression, a difference that could be of functional significance. In particular, the variability in somatic expression between neighbouring cells and the absolute somatic mean expression decreases along the axis, albeit quite weakly. There is no evidence of an interaction between the groups (see Fig. 6.6). This gradient is also present in Arc mRNA, as revealed in the Allen Brain Atlas.

This gradient could be functionally relevant because it corresponds with a gradient in the spatial resolution of grid cells in MEC (Hafting et al., 2005), as well as the synaptic integrative properties of stellate cells (Garden et al., 2008). I speculated in the previous chapter that dorsal cells could show higher levels of activity or an increased capacity for plasticity than ventral cells, but this requires further experimentation.

Dorsal to the MEC superficial layers are large patches of neurons that are not universally defined (cf. Boccara et al., 2010; Burgalossi et al., 2011). Arc-GFP activation during novel exploration does not distinguish these cells from layer II neurons, as they also show no clear increase.

6.5.3 *Novel exploration does not cause a detectable increase in Arc -GFP activation in the hippocampus*

The second surprising result from this experiment is that cells in the CA1 region do not show increased *Arc* expression following novel exploration. This is both surprising based on our understanding of CA1 and Arc, and it contrasts with previous work. CA1 has been an extensively studied region since it was discovered in the 1970s that CA1 pyramidal cells can act as place cells, modulating their firing depending on an animal's location within its environment (O'Keefe, 1976). Place cells make up approximately 25% of cells in CA1 at any one time (calculated from (Gothard et al., 1996b)) and can be detected using tetrode recordings when an animal explores a novel environment (Guzowski et al., 2006). Thus, it is highly likely that CA1 neurons were active in animals used in this experiment.

It is possible that the reason for these unexpected findings lies in the setup of this experiment. Previous work measuring Arc in CA1 has also used novel environments to up-regulate Arc in rats and mice (Guzowski et al., 1999; Christensen et al., 2013). In one example, rats were exposed to several novel environments before changes in *Arc* RNA expression were analysed (Guzowski et al., 1999). This work used cellular compartment analysis of temporal activity by fluorescent in situ hybridisation (catFISH) to show that the location of Arc RNA expression in CA1 can precisely reveal the history of activity of the neuron; recent activity is demonstrated by nuclear expression, whereas activity up to 30 minutes previously is evident from cytoplasmic expression (Guzowski et al., 1999). Further to this, Arc expression could be predicted using neurophysiological recordings of the same neurons (Guzowski et al., 1999).

There are several possible reasons for this discrepancy; task and pre-training (previous handling or not), or measurement (GFP expression vs FISH) could affect Arc expression. Of these, the mechanism of measurement is certainly a possibility. Although the measures used here are sensitive to novelty-induced changes in other cells, it is possible that in CA1 pyramidal cells, trafficking, degradation and dispersal of GFP is processed differently than in other cells. This could also mean that peak CA1 expression occurs at a different time point not measured here. The behavioural environment is unlikely to have an effect as in neither case did animals have a particular task to perform, although the frequent handling of the rats during novel exploration in the other experiment (Guzowski et al., 1999) could have encouraged the type of activity that forms place fields in CA1 cells. It seems unlikely that the lack of prior handling in the current experiment would reduce novelty-induced Arc in CA1, since it would be expected that, if anything, this would increase rather than decrease stress levels, and therefore Arc expression (Ons et al., 2004). Research has also revealed that long-term place field stability in CA1 cells is related to the attentional demands of the task (Kentros et al., 2004). Given that these animals had no task to perform, this may have limited the activation of Arc.

I did not quantify the protein expression of Arc in CA1 because other work in our lab suggests that protein expression is predominantly restricted to the dendrites in pyramidal layer neurons and I could not measure this with confidence with the Arc immunofluorescence technique in use. It would certainly be interesting to establish whether protein expression and promoter activity are dissociated in CA1. High-sensitivity FISH experiments in wild-type mice, such as those performed by Guzowski et al. (1999), would be most appropriate for indicating stimulus-related transcriptional processes, while protein expression could be analysed using Western blots on dissected tissue from the CA1 region such that dendritic Arc could be quantified. In addition to this, catFISH experiments could be used to test for differences in the activation of CA1 cells to novel exploration and a non-spatial task, such as interaction with an unfamiliar animal, to assess whether Arc transcription shows behavioural specificity.

Although the result for CA1 is surprising, it is perhaps highly consistent with the results described previously from MEC layer II. CA1 receives inputs from CA3 via the Schaffer Collateral, which itself receives inputs from superficial MEC via the perforant path, and CA1 also receives direct inputs from superficial MEC (summarised in (Canto et al., 2008)). Thus, if Arc-GFP activation does not increase in MEC layer II, it may follow that it may also remain at similar levels in the dentate gyrus and CA1. However, it remains to be understood why Arc expression is low in these regions.

6.5.4 *Cells in MEC deep layers considerably increase levels of Arc -GFP activation and Arc protein expression when a mouse explores a new environment*

The most striking change in Arc expression observed was in the dorsal region of layer V/VI of MEC, which directly contrasts with the result found in MEC layer II, and in the hippocampus. This result is particularly surprising given recent data from juxtacellular recordings that shows that when naïve rats are placed into a novel environment, layer II cells fire in a spatially

modulated manner, whilst layer V cells remain relatively silent (Burgalossi et al., 2011). However, extracellular recordings in animals with prior spatial experience have shown grid-like activity and high levels of activity in layer V (Hafting et al., 2005; Sargolini et al., 2006). While Burgalossi et al. (2011) suggest that their result may be precisely because their animals are naïve, such that networks between MEC and hippocampus do not yet contribute to spatial processing, the data I present here directly challenges this interpretation. Rather, it is possible that residual effects of anaesthesia in the rats used by Burgalossi et al. (2011) could contribute to their low firing rate, which could have important implications for the data they present concerning other components of the network. One important observation is that spatial firing is typically less stable in completely novel environments (Hafting et al., 2005; Langston et al., 2010), which could affect recorded levels of activity.

Increased Arc expression in MEC deep layers could be stimulated by a number of different possible inputs. The deep layers of MEC not only receive anatomical projections from the CA1 region of the hippocampus (Swanson and Cowan, 1977) and the subiculum (Kloosterman et al., 2003), but also the retrosplenial cortex (Jones and Witter, 2007). The apical dendrites of layer V/VI cells extend into the superficial layers where they receive inputs from presubiculum (Canto et al., 2012), and may contact other axons arising from cells in multiple cortical regions. Inputs from CA1 and subiculum have been shown to be excitatory and plastic (Craig and Commins, 2007, 2006). Although it seems unlikely that layer V/VI Arc increases are driven by CA1 activity, since I found no changes in CA1 Arc-GFP activation across any group of animals, this continues to be considered the dominant input to the deep layers and may influence their activity. One recent experiment that examined both proximal and distal input to layer V cells suggests that distal inputs may actually be larger in amplitude than proximal inputs and can generate dendritic spikes that propagate to the soma (Medinilla et al., 2013). It is therefore possible that Arc expression in layer V/VI cells is stimulated via some of these inputs. However, since these inputs have not been fully characterised and plasticity mechanisms have not been reported thus far, it is unclear what role Arc might play in these cells. The electrophysiological profile of many layer V cells supports a role for them in working memory (Egorov et al., 2002), so changes in Arc expression could reflect a demand for cells to modulate their synaptic strength.

An experiment that could clarify the inputs that Arc expression is stimulated by is to identify the types of neurons that do increase Arc and those that don't. The deep layers are populated by a range of neurons (summarised in Canto et al. (2008)) that differ in their morphology and dendritic distribution, and therefore the inputs they are likely to receive, and in their intrinsic properties (Hamam et al., 2000). Currently, molecular descriptions of these cells that relate to the properties described and that could enable selective targeting are limited. It is possible that some of the genes identified as showing selective expression within the deep layers in Chapter 4 could help elucidate this. An experimental method of identifying these cells would be to fill exploration-induced GFP expressing cells, either in acute slices or in fixed tissues, using a fluorescent polar dye so that morphology could be assessed. To investigate the electrophysiological properties, GFP-expressing neurons could be recorded from using whole-cell patch-clamp recordings. These methods would provide insight into the cell type and physiology, and the positioning of the cell's dendrites and axons,

thereby enabling an assessment of where the cell receives its inputs from. Paired recordings and optogenetic experiments could be used to ascertain functional inputs.

6.5.5 *Arc expression increases in the visual cortex following novel exploration*

In the visual cortex, *Arc* -GFP activation shows detectable but less robust increases following exposure to a novel environment than MEC deep layers (see Figs. 6.8&6.14). In layer VI, the deep output layer, the number of moderately-expressing cells that can be detected increases 3-fold following novelty compared to at baseline (Fig. 6.14). In the superficial layers Arc-GFP activation is higher following novelty using some measures, but this effect is not so consistent across all pairs as in MEC deep layers (Fig. 6.13). One possible reason for the less clear result in visual cortex is that mice continue to receive visual inputs under baseline conditions, which may be sufficient to up-regulate *Arc* -GFP activation in some cells. Further to this, I did not control for whether naive home-cage control animals were asleep or awake at the time of sacrifice, which could affect activity in visual cortex. Researchers who have used the same mice showed much lower Arc-GFP activation, in visual cortex when animals were maintained in the dark (Wang et al., 2006), so the fact that animals used here were sacrificed under light conditions could have reduced the size of the effect here.

An effect that can be observed here but that I have not quantified is that primary dendrites labelled with GFP appear to also be more easily detectable in novel-environment tissue. Novel exploration may also increase dendritic GFP labelling in background tissue, which is a limitation of the quantification method used here.

There is no clear pattern distinguishing *Arc* up-regulation in deep and superficial layers of visual and MEC. In the MEC novelty increases Arc-GFP activation more consistently and to higher levels in deep layers than superficial layers, but in visual cortex, the pattern of changes is very similar across superficial and deep layers, although the change is perhaps more pronounced and consistent in deep layers.

6.5.6 *The use of Arc as a measure of exploration-induced activity*

Arc is commonly used as a measure of neuronal activity. My aim in this chapter has been to use Arc as an indicator of cells throughout the MEC circuit that are both active and potentially undergoing plasticity-related changes. Although Arc and other immediate early genes are all used as reporters of neuronal activity, the reason that Arc is particularly useful for this is that research previously has dissociated Arc transcription from neural activity under specific conditions (Guzowski et al., 2006; Miyashita et al., 2009), one possible interpretation being that its expression is modulated by behavioural importance. However, since the extent to which Arc reports activity or relevance is not clear, the interpretation of Arc expression might be enhanced by measuring gene expression of genes such as c-fos or zif268, which are transcription factors that are not dendritically trafficked. A gene that more directly reports activity could clarify the activity levels of layer II MEC and CA1 cells at baseline and during novel exploration. These results could be contrasted with those of Arc to provide an indication of differences between activity and plasticity-related mechanisms.

6.5.7 *The use of Arc-GFP mice to measure exploration-induced activity*

As discussed in the previous chapter, there are caveats to the use of heterozygous Arc-GFP mice to report on Arc expression. These mice do not express wild-type levels of Arc protein, there are unknowns regarding the extent to which the translation of GFP will be affected by regulatory proteins that influence the translation of Arc protein, which limits the ability of GFP to report on Arc promoter activity, and the transcription of GFP RNA may be influenced by the strong promoter activity of the neo cassette in the transgene. However, these mice potentially provide two complementary methods of measuring Arc activity. This overcomes the limitation of only measuring mRNA, which does not provide information about the level of functional protein, or only measuring protein, which does not necessarily directly report Arc promoter activation because of trafficking and post-translational regulation. An important question is whether these potential benefits outweigh the limitations.

Arc protein plays an important role in synaptic mechanisms, including the internalisation of AMPA receptors (Chowdhury et al., 2006). There is also evidence that *Arc* protein, despite being a cytoskeletal protein, could influence its own and other protein levels through its effects on the ability of RNA to dock at synapses and on the phosphorylation state of other proteins (reviewed in Bramham et al., 2010). This leads to some ambiguity regarding the activity-dependent regulation of *Arc* in animals that possess reduced levels of basal *Arc* protein. One way to reduce this ambiguity would be to show in wild-type animals that the same patterns of activation of Arc protein exist. Another possible consequence of reduced Arc protein is altered hippocampal and MEC function. This has not been explicitly tested in heterozygous Arc-GFP mice but the lack of behavioural abnormality (Peebles et al., 2010) and evidence that Arc is not highly expressed during early development (ABA) suggests that these circuits are unlikely to be substantially altered.

Despite their reduced levels of Arc protein, I have used heterozygous Arc-GFP mice to report principally on the distribution of promoter activity of Arc across cells. Homozygous mice possess no detectable RNA corresponding to the C′ terminus of the Arc protein (Wang et al., 2006), and therefore GFP RNA is unlikely to be trafficked or degraded as Arc RNA would be due to regions in its 3′ UTR (Kobayashi et al., 2005). As such, levels of somatic GFP RNA are likely to reflect Arc promoter activity. However, a protein must be translated in order to view Arc and GFP RNA presumably contains the same 5′UTR as Arc RNA, which could influence its translation. It has been shown that regulatory proteins can suppress translation of Arc protein, although in the case of Fmrp, this appears to only reduce Arc protein in the dendrites (Niere et al., 2012). Since the effects of this in different cell types in vivo remains somewhat ambiguous, it could be argued that GFP protein does not faithfully report on promoter activity, but indicates cells in which both promoter activity and successful translation occur. GFP could also fail to indicate promoter activity because its transcription could be negatively influenced by the strong promoter activity of the neo cassette (c.f. Olson et al., 1996). Removal of this neo cassette would have prevented this problem. This is an important limitation of these mice, but it remains likely that GFP expression provides a close reporter of cells expressing Arc, and remains a useful complementary measure to Arc protein, which shows

similar patterns of changes to it. Importantly, previous work has shown that the degradation of d2EGFP is not affected by neural activity (Sutton, as cited in (Wang et al., 2006)).

Despite the potential caveats, the advantage of using Arc-GFP mice is that it may be possible to consider differences between the promoter activation of Arc and protein expression following novel exploration. I found that changes in *Arc*-GFP activation and protein expression show broadly similar patterns, with increases in Arc-GFP activation always being matched by an increase in Arc protein where measured. This is consistent with the data presented by Wang et al. (2006) of double immunolabelling in visual cortex where the majority of cells show overlapping expression patterns. However, two observations may provide insight into the behavioural regulation of Arc. First, in MEC layer II somatic *Arc* protein increases following novel exploration whereas there is no evidence of any such increase in Arc-GFP activation in Arc^{+}/GFP^{+} mice. The second observation is that in GFP^{+}/GFP^{+} mice Arc-GFP activation increases following novel exploration within MEC layer II neurons, particularly in dorsal regions. It is possible that this simply results from increased levels of GFP in GFP^{+}/GFP^{+} mice, which could make the change in Arc-GFP activation appear greater following novelty in GFP^{+}/GFP^{+} mice than Arc^{+}/GFP^{+} mice. It is also possible that these effects are artefacts of the Arc-GFP expression. An alternative explanation is that the increase in Arc-GFP activation in dorsal cells following novelty occurs precisely because Arc protein is absent. It has been shown in a line with a null mutation following Cre-mediated type 1 recombination on the floxed KO gene locus that lack of Arc enhances early-phase LTP, disrupts late-phase LTP, and causes corresponding deficits in spatial learning (Plath et al., 2006). Enhanced early LTP could lead to increased transcription of immediate early genes. One possible explanation that may account for both observations is that activity-dependent increases in Arc protein reduce the subsequent potential for plasticity of synapses, thereby reducing further plasticity-related transcription or affecting levels of translation of other proteins. It is noteworthy that somatic expression in the cells of GFP^{+}/GFP^{+} mice appears to show a more pronounced reduction along the gradient, particularly following novel exploration (Fig. 6.17), which suggests that the difference between the genotypes is not simply increased GFP RNA.

It is likely given that the regulation of Arc, and GFP in these mice, *in vivo* is not well understood that our conclusions could be more confident if confirmed with immunohistochemistry experiments and steady-state ISH experiments at multiple time points in wild-type animals and Arc-GFP animals. It is possible that protein expression, whether Arc or GFP, peaks at different time points in different cells in different regions, which could also explain the apparent lack of novelty induction in CA1. To view Arc activity using Arc-GFP mice requires both translation of GFP and sacrifice of the animal at an appropriate time. Guzowski et al. (1999) showed using FISH that the window of Arc expression in CA1 is very narrow in that Arc RNA is not detectable in neuronal somata 1 hour after exposure to an environment. It is therefore possible that GFP expression was analysed too late. Although experiments would likely have to be performed in different animals given that RNA and protein expression peaks at different times (cf Guzowski et al., 2001; Plath et al., 2006; Wallace et al., 1998), the results could reveal how closely the distribution of GFP expression across cells matches Arc RNA expression and protein expression at different time points and across different brain regions. It is common for researchers to use in situ Arc RNA expression rather than protein as a

measure, so performing steady-state ISH experiments would allow close comparison with the work of others. High-sensitivity FISH experiments in which RNA can be measured in the nucleus within minutes of stimulus exposure would provide the most direct indication that novel exploration increases or does not increase Arc transcription.

6.5.8 *What does novel exploration stimulate?*

An aspect that potentially limits the conclusions that can be drawn from this work is that we do not know what aspect of the novel environment experience is causing *Arc* up-regulation, e.g. stress, novelty, movement, olfactory stimulation, visual stimulation. This could be particularly relevant for understanding the interactions between group (home-cage vs. novel environment) and region (MEC vs. hippocampus vs. visual cortex) in that different behavioural situations could create a different pattern of novelty-induced changes or lack of changes. However, the behavioural experience provided to the mice is arguably a common natural occurrence in mammals and elucidating the expression of an immediate early gene in spatial navigation circuits provides insight into the spatial encoding of novel environments.

6.5.9 *Possible methodological improvements*

An important question is whether this immunofluorescence method, followed by somatic quantification, was simply not sensitive enough to detect regional changes that are actually present, for example in CA1, MEC and the large patches. Certainly, extracellular, juxtacellular and whole cell recordings *in vivo* all show that layer II cells are active during exploration (Hafting et al., 2005; Burgalossi et al., 2011; Domnisoru et al., 2013; Schmidt-Hieber and Häusser, 2013), but it is possible that they do not require plasticity-related genes. I could have used alternative methods to immunofluorescence that provided increased amplification or a clearer distinction between signal and background to test this.

A limitation of the analysis used here is the limited number of sections used for quantifying expression in each region. I noted particularly in regions where GFP expression was low, such as MEC layer II, that there could be quite large differences between GFP antibody expression and direct GFP imaging, possibly due to sparse, non-uniform expression. Given such sparse expression, it would be preferable to average across a larger number of sections to clarify the result.

6.5.10 *Open questions*

The increase of Arc in MEC layer V/VI is a highly robust finding. However, this type of analysis was not ideal for investigating the presence of a gradient in MEC layer II cells. Does a gradient exist in Arc RNA and protein levels, does it exist across all mice, and does it impact on cell function? It would also be interesting to consider the wider activation of Arc during novel exploration, particularly in regions of the spatial navigation circuit such as the

retrosplenial cortex and subiculum, which are closely associated with the MEC deep layers and CA1.

6.5.11 *Looking ahead: does hippocampal input drive up-regulation of Arc in MEC deep layers?*

The most clear and robust result of this experiment is that novel exploration significantly increases Arc-GFP activation and protein expression in layer V/VI of MEC. The next step is to ask what the requirements are for increased Arc expression. What inputs does layer V/VI receive that either cause high cellular activity, or drive a specific requirement for cells to modulate their plasticity using proteins such as Arc? The primary input to dorsal MEC deep layers is considered to be via dorsal proximal CA1, which does not show activity-dependent Arc-GFP activation. In the next chapter, I investigate whether novel exploration-induced up-regulation of Arc expression depends on this hippocampal input.

7

THE EFFECT OF HIPPOCAMPAL LESIONS ON ARC EXPRESSION IN MEC

7.1 SUMMARY

I showed in the previous chapter that novel exploration significantly up-regulates GFP expression in an Arc-GFP knockin mouse line in the layers of MEC that receive inputs from CA1, the deep layers. However, it has no detectable effect on levels of Arc-GFP activation in the superficial layer II of MEC, a region known to possess spatially modulated cells that fire when an animal is placed into a novel environment. It also has no detectable effect in dorsal CA1 pyramidal cells, which are known to form place fields in novel environments. These results led me to hypothesise that the dorsal CA1 input to dorsal MEC layer V/VI is not necessary for Arc up-regulation in MEC layer V/VI. Here, I show the following:

1. Lesions of dorsal CA1, which impair spatial working and reference memory, do not prevent up-regulation of *Arc* in MEC deep layers following novel exploration. This suggests that the information important to potential plasticity-related coding in this part of the circuit is not dependent on the major hippocampal output pathway from CA1.
2. Lesions of dorsal CA1 also do not prevent up-regulation of Arc-GFP activation in visual cortex. This suggests that CA1 is not necessary for stimulating the signal that distinguishes baseline visual experience from the experience of novel exploration.

7.2 INTRODUCTION

The entorhinal-hippocampal circuit is believed to be fundamental to the mechanism by which the brain encodes spatial environments. When an animal is placed into an environment grid cells in the MEC form a metric representation of space (Hafting et al., 2005). This activity is hypothesised to underlie the development of CA1 place fields (McNaughton et al., 2006a) that each encode a specific place (O'Keefe and Dostrovsky, 1971). This information is relayed back through the MEC (Swanson and Cowan, 1977) and out to other regions of neocortex via the deep layers (Insausti et al., 1997). Disruption of this circuit causes a variety of deficits in space-related learning, memory and spatial navigation (Morris et al., 1982; Moser et al., 1993).

Despite the potential significance of exploration-induced activity throughout the network to spatial navigation, the expression of a certain activity-regulated gene is only selectively increased in mice, occurring predominantly in the MEC region that sends axons to the neocortex (Insausti et al., 1997), layers V/VI (Section 6.5.4). The MEC layer that contains well-characterised grid cells and the CA1 place cells do not show detectable increases in Arc-GFP activation in a reporter mouse line in which GFP is under the control of the Arc promoter, despite typically showing activity in the type of novel exploration experienced. This raises two questions: (1) do deep layer neurons encode important information during novel exploration?, and (2) is this process dependent on information from the hippocampal formation which provides a major input (Swanson and Cowan, 1977).

Extensive and detailed anatomical work has characterised the anatomical connectivity between the hippocampus and MEC. MEC layer II projects topographically to the DG, CA3 and CA1 (Dolorfo and Amaral, 1998b; van Groen et al., 2003), whereas deep layers receives parallel topographical projections from CA1 and subiculum (Swanson and Cowan, 1977; Tamamaki and Nojyo, 1995; Kloosterman et al., 2003; Cenquizca and Swanson, 2007) (Fig. 7.1(A)). Retrograde injections into regions of MEC that are close to the rhinal sulcus, which in MEC are the dorsal regions, strongly label septal and proximal CA1 and more weakly label septal, distal subiculum (Tamamaki and Nojyo, 1995; Kloosterman et al., 2003) (See Fig. 7.1(B)). Thus, dorsal MEC deep layers, which show activity-dependent changes in Arc expression, likely receive a major input from proximal, septal CA1.

The function of this input is not well understood. Hippocampal inputs to MEC deep layers have the potential for plasticity (Craig and Commins, 2006, 2007), but this has not been demonstrated *in vivo*. Inputs target layer V neurons at the proximal and basal dendrites (Medinilla et al., 2013). Proximal inputs, which could also include inputs from other regions (Jones and Witter, 2007), have been shown to be excitatory and predominantly AMPA-mediated (Medinilla et al., 2013). A hippocampal projection could therefore drive Arc expression in deep layer neurons.

Although the hippocampal input appears to be functional, other inputs to layer V/VI could be of greater significance to the information processing in cells that up-regulates levels of a plasticity-related protein during novel exploration. Layer V neurons have long, spiny apical dendrites that extend into the superficial layers, thereby providing them with the potential to receive inputs that terminate in any layer of MEC (Hamam et al., 2000; Medinilla et al., 2013). Indeed, layer V cells receive inputs from pre- and para-subiculum that target the superficial

layers (Canto et al., 2012). It is therefore possible that increases in Arc expression that follow novel exploration could be independent, or at least, rely minimally, on a functional input from the hippocampus, and instead depend on inputs from cortical regions that project axons through MEC (Insausti et al., 1997).

To test the possibility that novel exploration-induced up-regulation of Arc expression in MEC layer V/VI can occur independently of CA1, I lesioned the hippocampus using ibotenic acid lesions. These are excitotoxic and cause cell death, but do not damage fibres of passage or vasculature and have been used previously to isolate selective deficits (Jarrard, 1989). If ibotenic acid lesions prevent Arc up-regulation in MEC layer V/VI, this suggests that the hippocampal input is required, even if CA1 cells do not themselves express Arc during novel exploration. If lesions do not prevent up-regulation, this suggests that potential plasticity-related changes in layer V/VI cells may be coding information that is not directly related to hippocampally processed information.

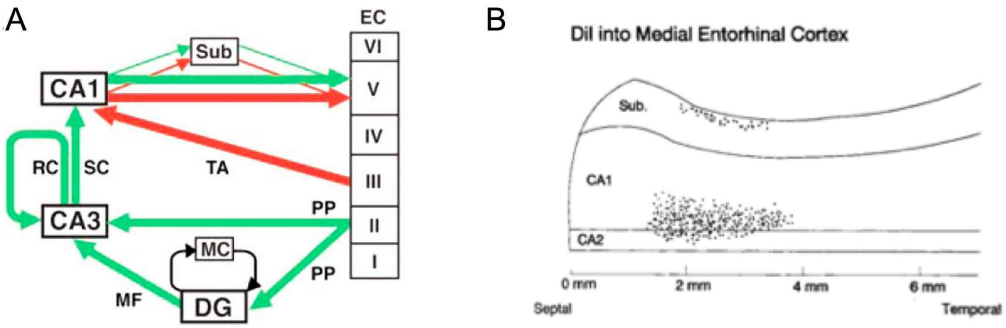


Figure 7.1: *Connectivity between the hippocampus and MEC* (A) Schematic (from Nakashiba et al. (2008)) showing the connectivity between the entorhinal cortex (EC) and hippocampal subfields. MEC layer V/VI receives a major input from CA1 and a smaller input from the subiculum, which itself receives inputs from CA1. (B) Schematic from (Tamamaki and Nojyo, 1995) showing the neurons in the septal half of proximal CA1 (black dots) that project to dorsal MEC, as shown using a retrograde lipophilic tracer, DiI, injected into dorsal MEC.

7.3 AIMS

1. Does novel exploration-induced Arc-GFP activation in MEC and visual deep layers depend on a functional CA1 input?
2. Is there a correlation between Arc-GFP activation in the left and right hemisphere? Because of the requirement to analyse lesion extent in both hemispheres, it follows that I should investigate how similar expression is in left and right.
3. Are lesion extent and increases in Arc-GFP activation correlated? Even if the CA1 input to MEC layers V/VI is not necessary for Arc-GFP activation to increase, it may contribute to it.

7.4 METHODS

7.4.1 *Animals used*

As in previous chapters, heterozygous Arc^{+}/GFP^{+} mice were used. These mice express a GFP construct under the control of the Arc promoter, thereby revealing the distribution of Arc-GFP activation at a cellular level *in vivo*. All 7 littermate pairs were aged 6-9 weeks at the time of surgery and between 7 and 11 weeks at the time of novel environment exposure and sacrifice. The mice used in this experiment were all bred from GFP^{+}/GFP^{+} (female) and Arc^{+}/Arc^{+} (male) pairs.

7.4.2 *Methods for hippocampal lesion surgery*

Hippocampal lesions were performed on each animal using ibotenic acid injected at multiple sites. The aim was to lesion the entire dorsal CA1 region, with additional targeting of intermediate regions to ensure that as many cells projecting to dorsal MEC deep layers were killed as possible. Lesions were not explicitly designed to avoid or target CA2, CA3 or DG. All surgical procedures were performed with sterile solutions and autoclaved, bead heat-treated or sterile equipment (see Fig. 7.2(A) for experimental setup). The level of anaesthesia was monitored and adjusted continually throughout the procedure depending on the breathing rate and size of each animal. All animals were injected with an analgesic and were post-operatively injected with ~1ml warmed saline to prevent dehydration.

For the surgery procedure each mouse was removed from its home cage and placed into a small anaesthesia chamber receiving 2% isofluorane in oxygen. Depth of anaesthesia was assessed by breathing rate. Once breathing had stabilised and no response was seen to gentle tilting of the chamber, each mouse was removed and placed into a stereotaxic frame with a tooth bar in which to insert the upper incisors. The tooth bar was connected to a tube through which 2% isofluorane could be passed. A toe pinch was used to check depth of anaesthesia before fixing the animal into the frame. Viscotears was placed onto the eyes to prevent corneal drying during surgery. The animal was secured by inserting blunt ear bars into the lower part of each ear, ensuring first that the upper surface of the skull was parallel to the bench surface (see Fig 7.2(A)). Bars were inserted until resistance could be felt and the skull showed no movement when gently tapped with a sterile cotton bud. Anaesthesia was maintained at 2% until the animal was securely fixed into the frame and its breathing rate was steady, at which point it was reduced to 1.5-1.75% depending on animal size. This proportion was steadily decreased to ~0.75-1% isofluorane during the surgery.

An analgesic (0.03ml Vetergesic) was injected subcutaneously into the rump of skin on the back prior to beginning surgical procedures. Betadine was used to clean the area on the animal's head surrounding the wound. Fur was parted using curved forceps and the back of a sterile razor prior to making an incision with the razor that started behind the animal's eyes and finished behind its ears. The wound was opened and cleaned using sterile cotton buds and kept moist with warmed saline solution.

Bregma and lambda were identified by the position of the midline sagittal suture and blood vessel (see Fig. 7.2(B)). In cases where the connective tissue obscured the bone joins, a metal tool was used to scrape away the surface tissue. The x, y and z coordinates were determined using a coordinate system tool attached to the stereotaxic frame. A Dremel hand drill was fixed into the stereotaxic frame and aligned with Bregma and lambda using manually controlled dials. This provided a readout of the coordinates. If Bregma and lambda were not within 0.2mm in the z plane (indicating that the brain was not flat), the height of the tooth bar was adjusted to flatten the skull, and bregma and lambda were recalculated.

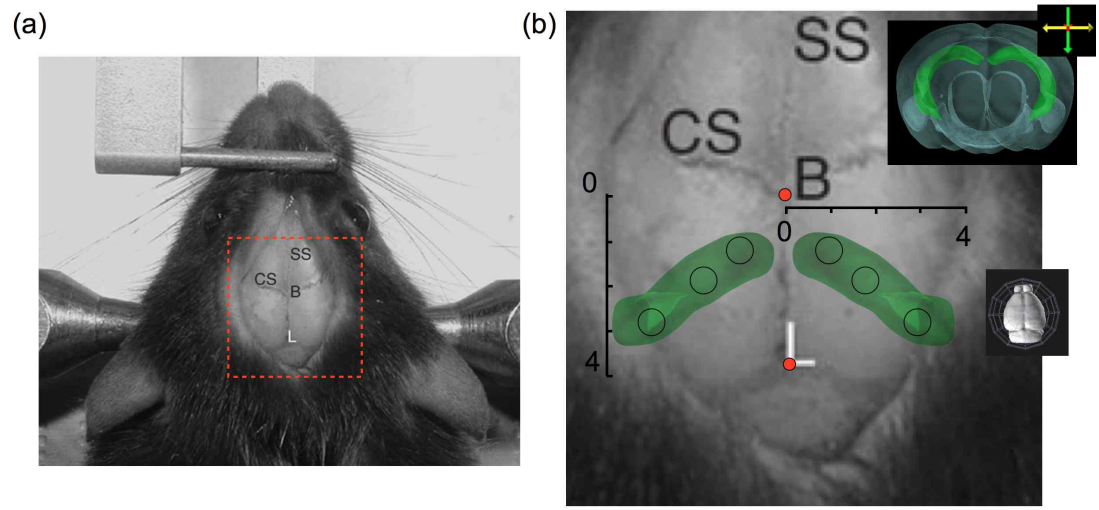


Figure 7.2: *Hippocampal lesion surgery* (a) Image showing the positioning of a mouse with blunt ears bars, a nose bar and a tooth bar to secure the head. The area outlined in red is shown in (b). (b) Image showing the positioning of Bregma (B) and Lambda (L) (red dots) when looking down onto the skull. Bregma sits at the junction of the sagittal suture (SS) and coronal suture (CS). Green areas (taken from the Allen Institute Brain Explorer II) show the CA1 region in 3D. Measurements indicate the 4mm posterior and lateral to Bregma. These measures were used to identify drill sites (black circles) that would target CA1. (inset) Image from the Allen Institute Brain Explorer II showing the 3D dorso-ventral extent of CA1 region from the posterior angle

Site	AP	ML	DV	Vol (nl)
1	1.6	1.1	1.8	100
2	2.2	2.2	1.8	100
3	3	3.1	2.7	100
4	3	3.1	4.5	100

Table 7.1: *Injection sites.* The targeted injection coordinates are shown, given a Bregma-lambda distance of 4.2mm. At each site 100nl was injected, except at site 3. These coordinates were adjusted based on B-L distance, if the brain was not completely horizontal, or if large blood vessels were located at the drill site.

For each injection site, the drill was moved to the coordinates of interest, using the coordinate tool as a readout. To minimise the possibility of haemorrhage I assessed the location of surface blood vessels by dampening the skull using saline prior to drilling so that it became transparent. Where blood vessels lay directly below the planned drill site, the coordinates were adjusted by up to 200 μm in 1-2 directions, depending on injection site. Once the coordinates were fixed and recorded, the skull was dried, the drill was switched on and the drill bit (0.7mm) slowly lowered onto the skull. The skull was drilled until cracks were evident and fluid began to seep through. At this point the drill was removed and a small needle with a very narrow tip was used to peel back and remove the remaining skull, then to pierce the dura. The brain underneath was kept moist with saline. The drill was then moved to the coordinates of the next injection site and the process repeated.

Once all drill sites had been made, a fine-tipped glass electrode hand pulled using a glass puller on the day of the surgery was attached to the stereotaxic frame. Ibotenic acid (2mg/ml in PBS) was taken up into the glass pipette using suction and it was aligned to bregma so that the coordinates could be recalculated. For each injection site, the pipette was lowered onto the surface of the brain and the reading from the coordinate system noted. All injections were made relative to bregma in the x-y plane and relative to the brain surface in z. The pipette was lowered through the dura and through the brain at a rate of 100 μm per second. At 200 μm beneath the desired injection point the pipette was retracted by 200 μm. After 30 seconds, 50 or 100nl of ibotenic acid was injected using pressure created by a syringe. This was calculated by the distance moved by the fluid in the pipette (1mm =100nl). After the injection, the pipette remained in place for 60 seconds before being slowly retracted. When more than 1 injection was made into the same drill hole, the most ventral injection was made first with shorter periods between injections. Table 7.1 shows the drill and injection sites and injection volumes. Injections were made symmetrically (where possible) in the two hemispheres. The extent of the lesion was confirmed post-mortem (see section 7.4.4).

Following injections the wound was closed using TissueTek tissue glue, the animal was removed from the stereotaxic frame and allowed to recover in a clean cage on a heat mat. Recovery was monitored and animals were returned to cages shared with others in the habituation room, once they were moving and feeding normally (typically within 24 hours). Animals that did not show a normal recovery (typically assessed after 12 hours) were euthanased. This occurred for one animal, possibly due to haemorrhage. Animals were kept in the habituation room for at least one week prior to novelty exposure.

7.4.3 *Behavioural procedure*

The procedure for exposure to a novel environment, tissue processing and image analysis was as described in the previous two chapters. Briefly, following at least 1 week of habituation to a quiet room, a pair of animals was removed from a cage. One 'home-cage' animal was directly anaesthetised and perfused while the other 'novel-environment' animal was placed into a novel arena for 15 minutes then returned to a clean cage for 90 minutes before perfusion. The movement of all animals in the maze was observed to ensure that animals were exploring the whole environment.

After sacrifice brains were fixed, cryoprotected, cut and imaged. Arc expression was measured using native GFP fluorescence and GFP antibody labelling. Differences in Arc expression were calculated based on labelling within neuronal somata relative to the surrounding region.

7.4.4 *Analysis of lesion extent*

7.4.4.1 *Tissue preparation*

Following perfusion, brains were post-fixed with 4% PFA for 20-24 hours at 4°then cryoprotected in 30% sucrose. They were frozen and cut to 40µm in the sagittal plane using a freezing microtome. Sections containing the hippocampus at 240 µm intervals were mounted on Superfrost slides that electrostatically adhere tissue sections to the glass. Sections were air-dried then dehydrated through a series of ethanol solutions. After rinsing in water, sections were stained with cresyl violet solution for 15 minutes before further dehydration in ethanol. Sections were cleared with xylene then mounted in Eukitt (See Appendix A for the full protocol).

7.4.4.2 *Imaging Nissl-stained tissue*

Sections were imaged at 4x magnification with brightfield illumination using a CCD camera. Approximately 20 images were collected of CA1 throughout the medio-lateral extent for dorsal through ventral regions (See Fig 7.3(A)). Images of Nissl stained hippocampus were manually aligned to atlas images so that the border of CA2 and CA1 could be more precisely estimated. Dorsal hippocampus was defined as between 0 and 2.6mm medio-laterally, intermediate hippocampus as from 2.6mm laterally down to 3.5mm ventrally. CA1 regions more ventral than this were considered ventral hippocampus. Fig. 7.3(B) shows these divisions, which were calculated based on divisions in rat hippocampus made by Jarrard et al. (2012), experiments in mice by Murray et al. (2011) that disrupted spatial reference and working memory, and using data from retrograde tracing experiments performed in my lab.

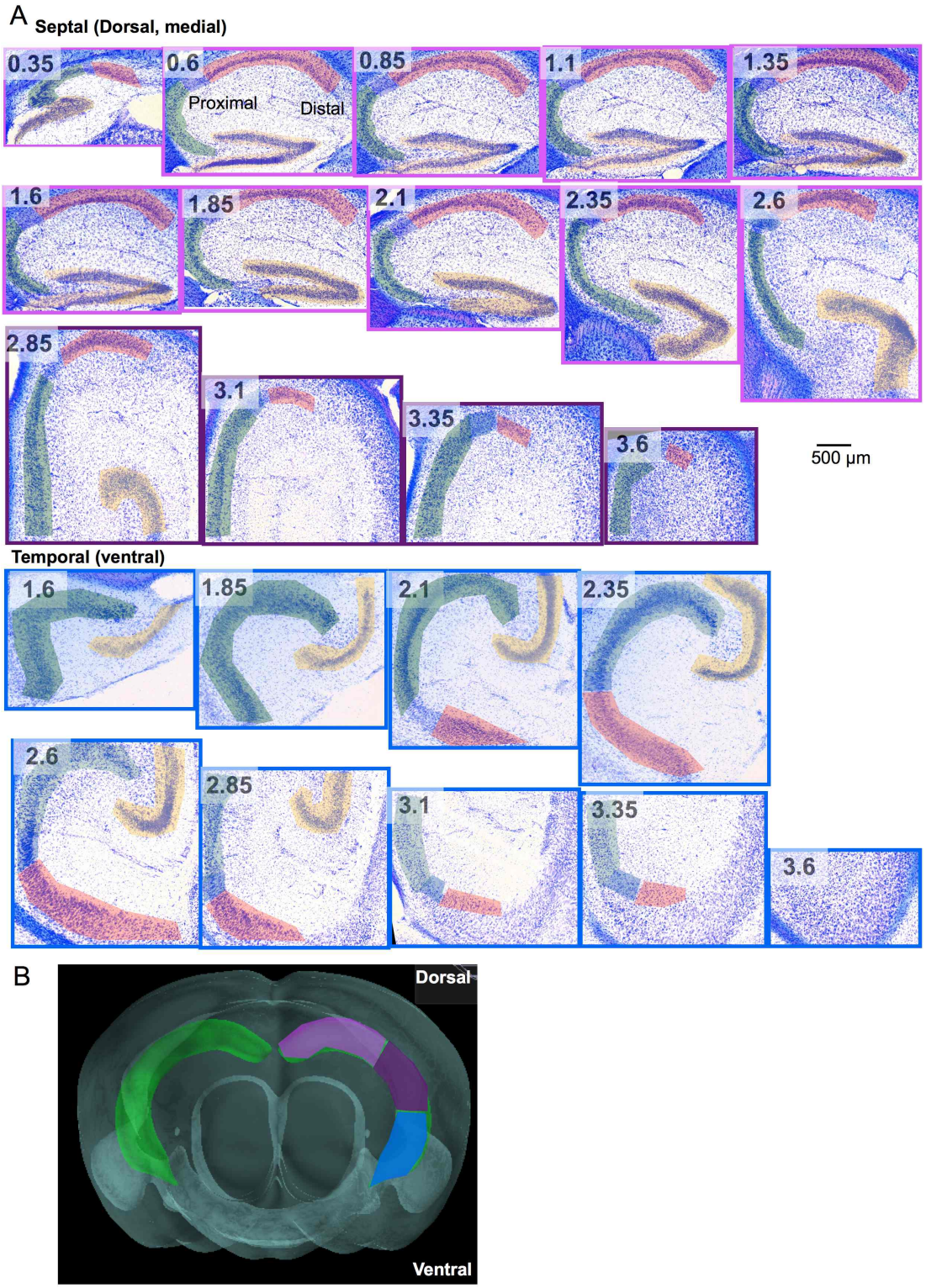


Figure 7.3: *Sections used for analysis of lesion extent.* (A) Nissl-stained sagittal sections of the hippocampus throughout the medio-lateral extent from septal (upper left) to temporal (lower). Numbers indicate the distance from the midline in mm. Images outlined in magenta were defined as 'dorsal', in purple as 'intermediate', and in blue as 'ventral'. The CA1 region is marked in red, CA2 in blue, CA3 in green and dentate gyrus in yellow. (B) Image from the Allen Brain Explorer II overlaid with the divisions of the CA1 region into dorsal, intermediate and ventral regions.

7.4.4.3 *Analysis of Nissl-stained tissue*

Damaged tissue was defined based on the absence of large brightly stained pyramidal cell bodies that could be clearly distinguished from the molecular layer of CA1 (Fig. 7.4(A)), using images from Jarrard (1989) (Fig. 7.4(B)). Assessment of damaged tissue was made blind to the experimental condition and to the results of the Arc-GFP activation analysis.

The analysis was performed using reference atlas sections from the Allen Brain Atlas as a measure so that all lesion data could be placed into the same framework. Prior to assessment of tissue sections, the pyramidal layers of CA1, CA2 and CA3 and the granule layer of the DG were outlined on each reference atlas section using the software GIMP. This generated one 2D binary image for each region within each section (Fig. 7.4(C)(a)). These 2D binary images could be categorised as dorsal, intermediate or ventral, as described previously.

For each mouse tissue section the nearest sagittal reference atlas section was chosen and coarsely aligned to the tissue section (Fig. 7.4(D)). The entire damaged area for that section was outlined using the GIMP path tool onto the corresponding reference atlas (Fig. 7.4(C)(b)). To aid in estimating the extent of damage within each region so that it could be outlined, the length of the full cell body region and damaged region were measured and compared using manually drawn Bezier lines (Fig. 7.4(D)). The area of damage to the cell body layers within each region on each section could be calculated by multiplying the 'damage' image with the binary atlas image (Fig. 7.4(C)(c)). This analysis provided both the number of damaged pixels as a percentage of the total number of pixels corresponding to each atlas region and an image representing the damage to each region. To calculate the total amount of damage per region for each mouse and each hemisphere, the damage was summed across all the sections containing that region and a percentage of the total calculated. To calculate total amount of damage across CA1, CA3 and the DG, the percentage damage to each sub-region was calculated and weighted by the total length of the cell body layer measured across multiple sample sections. To visualise the lesions, the images representing damage to cell body layers for all mice were overlaid and colour-coded based on the consistency of the lesion.

Hemispheres with less than 60% of dorsal CA1 damaged were not included in the analysis of comparisons between home-cage and novel-environment animals. This measure was used by (Murray et al., 2011), who showed impairments in spatial working and reference memory in mice with lesions greater than this size. For all GFP antibody results, data from each hemisphere are presented, followed by data from the hemisphere with a greater degree of damage. For direct GFP fluorescence, images were all acquired from the left hemisphere. In animals in which the left hemisphere was not sufficiently lesioned, direct GFP fluorescence data were not included.

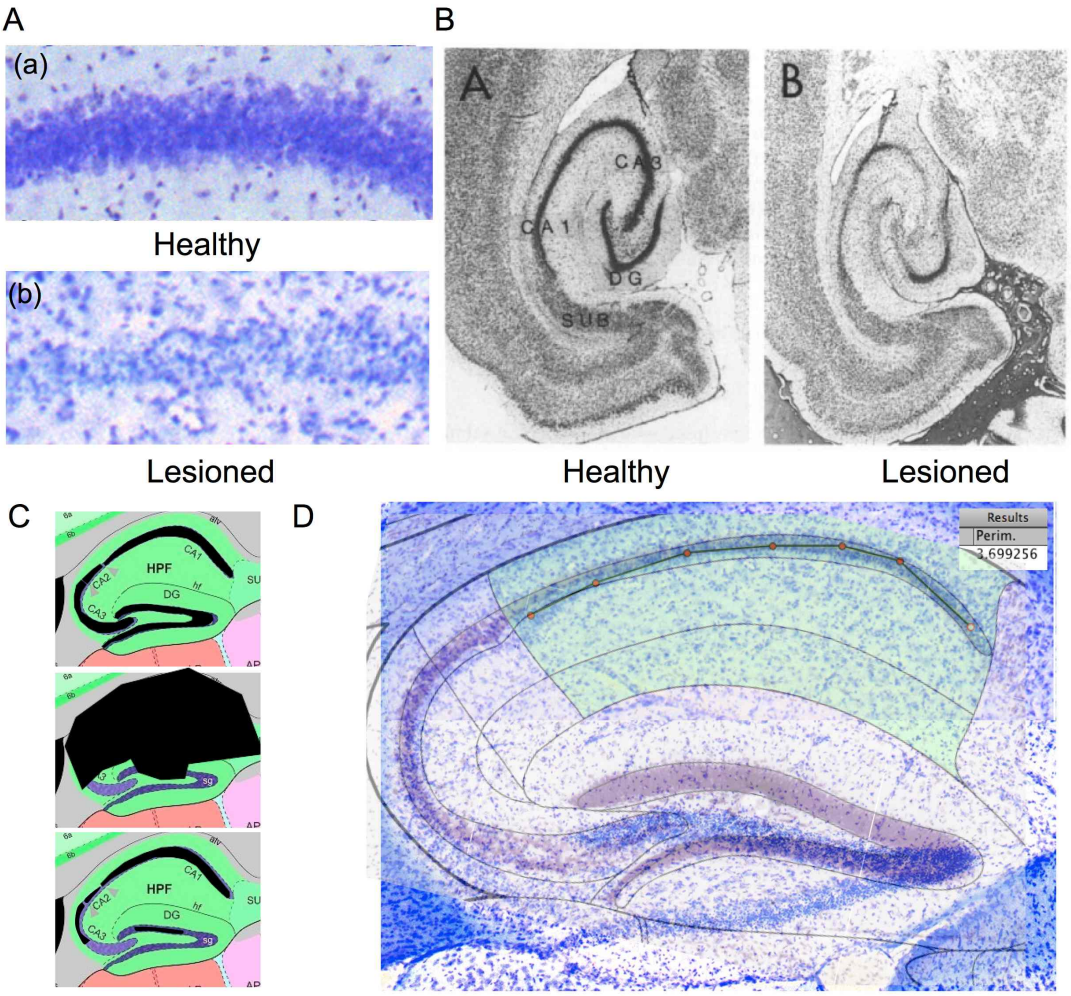


Figure 7.4: *Analysing lesion extent* (A) Images show examples of the CA1 pyramidal layer in (a) healthy and (b) damaged Nissl-stained tissue. Healthy tissue is characterised by the presence of densely packed brightly stained patches with a cellular appearance whereas damaged tissue contains no clear cell structures and predominantly appears to be debris or non-neuronal cells. (B) Image of horizontal Nissl-stained sections from Jarrard (1989) to show the difference between healthy and lesioned tissue in animals sacrificed 4 days post-surgery. (C) Procedure for estimating lesion extent:(a) Regions were defined using the Allen Brain Atlas reference atlas. (b) For each mouse and each hemisphere, the damaged area was outlined. (c) The resulting damage for each region could be calculated from the two images. (D) Example Nissl-stained section overlaid with Allen Brain Atlas reference atlas divisions. The faint green area indicates CA1. The black line with red dots indicates the line drawn over the CA1 pyramidal layer to assess the proportion of damage. (inset) Results box shows the length of the manually drawn line.

7.4.5 *Region analysis*

In this chapter I have focussed primarily on changes in the MEC deep layers, with visual cortex deep layers as a secondary comparison on interest. I have also directly compared the left and right hemispheres on MEC on the basis that CA1 output to MEC is unilateral (Swanson et al., 1978).

7.4.6 *Data quantification, presentation and statistics.*

Methods for data quantification were predominantly as described in previous chapters. As in the previous chapters I have set thresholds for antibody labelling that help establish how exactly the distributions differ. Cells that exceed a 'detectable' threshold are those that are automatically detected for direct fluorescence, and neurons that have a less than 5% chance of being detected in control (secondary only) conditions for the GFP antibody method.

For direct GFP fluorescence 'moderate' cells have a soma-surround score in the top 50% of the range of scores for a given region (MEC deep layers or visual cortex deep layers), whereas 'high' cells have a soma-surround score in the top 25%. For GFP antibody labelling 'moderate' cells have a soma-surround score in the top 50% of the range of scores, whereas 'high' cells have a soma-surround score in the top 10%.

This experiment was performed in two sessions separated by a two month gap. Animals within each stage were matched on injection site, conditions in which they were held prior to the experiment, and all tissue processing stages, including microscopy session. However, across sessions there were differences in immunofluorescence and imaging conditions. I therefore performed the normalisation procedure described in Section 6.3.2 of the previous chapter. Briefly, all data points for all animals for a given region were normalised to the session mean. Data points were further normalised to the range [0,1]. All tests for the effect of novel exploration were planned in advance of the experiment and have therefore not been corrected for multiple comparisons. I have not tested any unplanned effects.

7.5 RESULTS

I demonstrated in the previous chapter that Arc-GFP activation and Arc protein are up-regulated in MEC layer V/VI following exposure of an animal to a novel environment. A major projection to MEC layer V/VI is believed to arise from dorsal CA1 pyramidal cells. To test whether this input is necessary for changes in Arc in layer V, I lesioned the hippocampus and tested whether Arc continued to be up-regulated in MEC layer V/VI in the absence of a hippocampal input. All lesions occurred prior to animals having any experience of the novel environment, including the arena and the room containing it.

7.5.1 *Hippocampal lesion extent*

All animals received bilateral injections of ibotenic acid that were targeted to excitotoxically lesion the majority of the hippocampus, and in particular, dorsal CA1. Data were removed from the analysis on the basis of damage to the dorsal CA1 region. The dorsal CA1 region was consistently damaged, with at least 80% of cells lost in the pyramidal layer, on average, across left and right hemispheres (Fig. 7.5(A-B)). In almost half the animals no healthy cells could be detected in dorsal CA1 in at least one hemisphere. There were no differences in lesion extent between the home-cage and novel-environment groups (t-test: LHS $p = 0.63$, RHS $p = 0.15$). Nevertheless, lesions were not sufficient in the left hemisphere of one home-cage mouse and one novel-environment mouse and in the right hemisphere of one home-cage mouse. As such, comparisons between home-cage and novel-environment mice are made on the basis of 6 home-cage mice vs. 6 novel-environment mice for the left hemisphere and 6 home-cage mice vs. 7 novel-environment mice for the right hemisphere.

The full extent of cell loss throughout the hippocampus is indicated in (Fig. 7.5). The extent of cell loss throughout different parts of the hippocampus was more variable between mice than just the dorsal CA1 region. In particular, the ventral, medial aspect was rarely damaged, which is consistent with the fact that no injections were targeted to the most ventro-medial aspect.

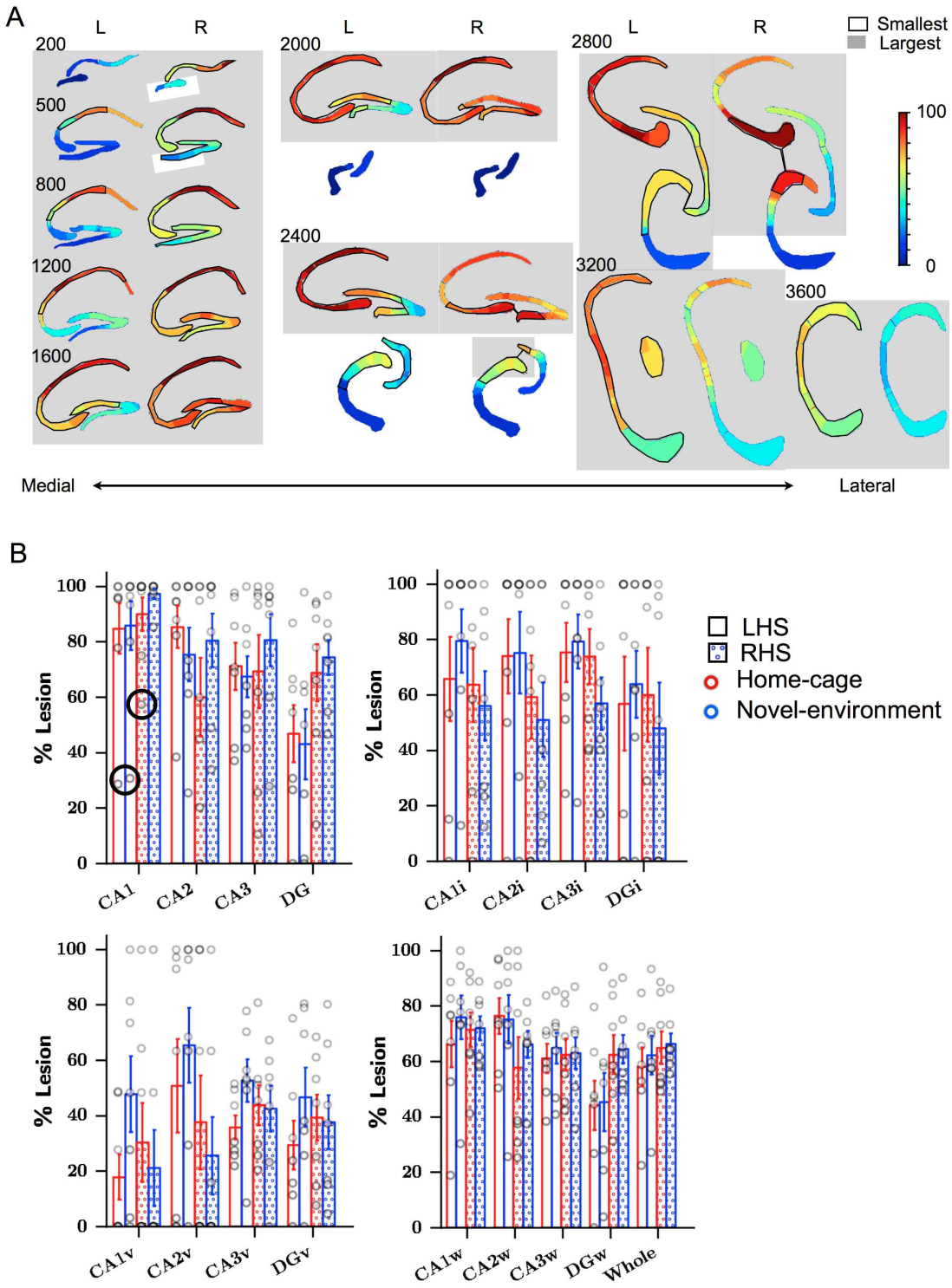


Figure 7.5: *Lesion extent throughout the hippocampus.* (A) Average images showing the consistency of lesions throughout all mice in the left hemisphere (left) and right hemisphere (right). Dark red areas are lesioned in all animals whereas dark blue areas have not been lesioned. Numbers indicate distance in μm from the midline. The black outline and grey square indicate lesions for mice with the smallest and largest dorsal CA1 lesions on the left and right hemispheres that were included in the analysis, respectively. (B) The percentage lesioned for each group (home-cage: red, novel-environment: blue) and each hemisphere (LHS: blank, RHS: filled circles) is plotted across dorsal, intermediate, ventral and the whole of CA1, CA2, CA3 and DG. The damage to the entire hippocampus is also shown. Data from hemispheres corresponding to points marked with black circles for dorsal CA1 were not included in subsequent analyses comparing Arc-GFP expression.

7.5.2 *Lesions of dorsal CA1 do not prevent up-regulation of Arc-GFP activation in MEC deep layers*

I measured Arc-GFP activation in Arc+/GFP+ mice by measuring the number of cells that can be directly detected using direct GFP fluorescence. The number of detectable cells in MEC layer V/VI is higher following novel exploration, as shown in the example images (Fig.7.6(A)), distributions (Fig.7.6(B)) and summary data (Fig.7.6(C): t-test: $p = 0.011$). The number of cells showing moderate expression (t-test: $p = 0.0003$) and high expression (t-test: $p = 0.0033$) are also significantly higher following novel exploration. Not only are there an increased number of detected cells, but the detected cells appear to be more uniformly labelled (Fig.7.6(D)).

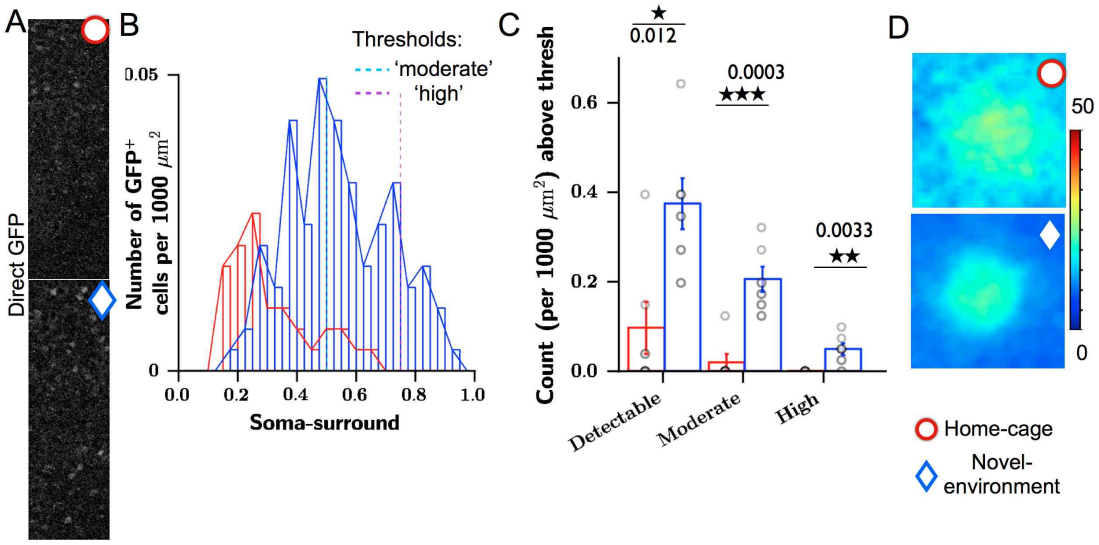


Figure 7.6: *Analysis of directly imaged GFP fluorescence in MEC layer V/VI in home-cage and novel-environment animals following CA1 lesions.* (A) Examples of confocal 20x maximum projection images from a home-cage (red circle) mouse and a novel-environment (blue triangle) mouse. (B) Histograms show the distributions of soma-surround score for cells detected using direct GFP fluorescence. Vertical lines mark the thresholds for the data plotted in (C). (C) The number of neurons exceeding thresholds marked in (B) is shown for home-cage and novel-environment animals. Results of a one-way ANOVA are shown, with alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. (D) Images show the average 60x60 pixel region of fluorescence surrounding a detected GFP^+ cell, calculated across all detected cells across all home-cage (red,*n=6*) and novel-environment (blue, *n=6*) animals that met inclusion criteria. All data taken from the left hemisphere.

I used GFP antibody labelling both to examine how neuronal expression of GFP differs following novel exploration and to test whether I could see different effects in the left and right hemispheres. As shown in Fig.7.7(A), GFP antibody labelling appears to increase in layer V/VI following novel exploration. In both left and right hemispheres, the distribution of soma-surround scores shifts to higher values (Fig.7.7(B)). This can be seen in the summary data (Fig.7.7(C)) in that both the mean soma-surround and standard deviation in soma-surround increase in the left hemisphere (t test - Mean: $p = 0.0003$, SD: $p < 0.0001$), the right hemisphere (t test - Mean: $p = 0.0001$, SD: $p < 0.0001$) and in the hemisphere with a greater lesion extent (t test - Mean: $p < 0.0001$, SD: $p < 0.0001$). The proportion of neurons with a moderate level of GFP labelling is also significantly higher following novel exploration in both hemispheres (t test - LHS: $p = 0.0002$, RHS: $p =< 0.0001$), as is the proportion of

high expressing neurons (t test - LHS: $p = 0.003$,RHS: $p = 0.0035$) (Fig.7.7(D)). Across mice, although the mean soma-surround scores varies, it is highly consistent across left and right hemisphere, although the data are not significantly correlated (Fig.7.7(E)). The similarity can be seen in the plots in Fig.7.7(F), which show an average cell in each hemisphere, at baseline, and following novel exploration.

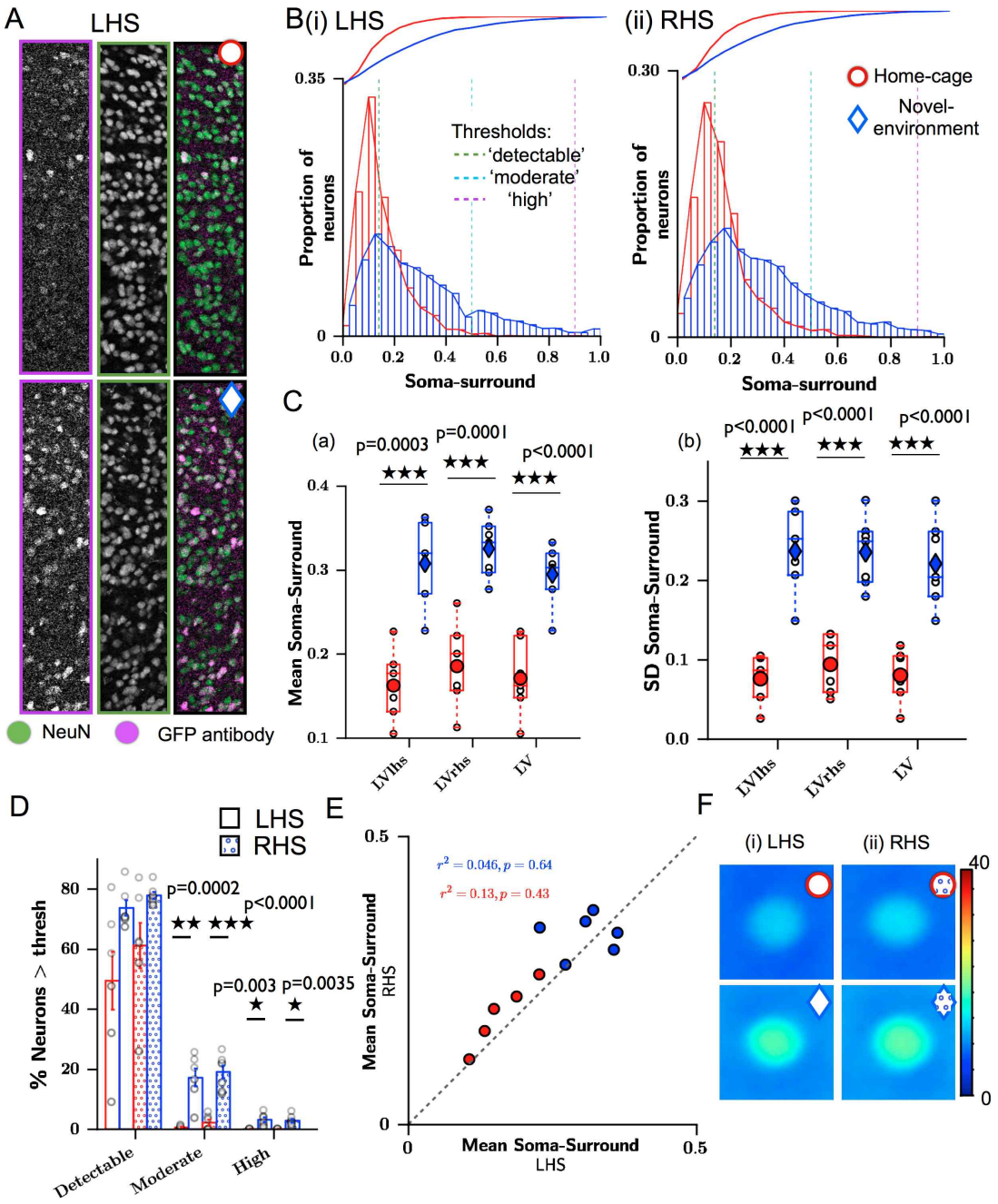


Figure 7.7: *Somatic GFP antibody labelling in MEC layer VI neurons is increased following novel-environment exposure in animals with CA1 lesions* (A) Examples of confocal 20x maximum projection images from a home-cage (red circle) and novel-environment (blue triangle). GFP antibody and NeuN labelling are shown outlined in magenta and green, respectively. The accompanying image shows the GFP/NeuN overlay. (B) Histograms show the distributions of soma-surround score for all neurons detected using NeuN labelling. Vertical lines mark the thresholds for the data plotted in (D). (C) Boxplots show (a) The mean GFP labelling soma-surround score for all neurons, and (b) the standard deviation in soma-surround score. The data plotted are for the left hemisphere (LVlhs), the right hemisphere (LVrhs) and the hemisphere with the greater amount of damage (LV). Filled shapes indicate the group mean. Open black circles indicate data for each mouse (HC-lhs: n=6, rhs:n=6; NE-lhs: n=6, rhs: n = 7). (D) The proportion of neurons exceeding GFP labelling thresholds marked in (B) is shown for mice in each group. Bars representing data from the LHS are shown blank whereas those for the RHS are filled. (E) The relationship between mean soma-surround scores is plotted for the LHS (x-axis) and RHS (y-axis) for mice in which both hemispheres have been sufficiently lesioned (HC:red: n=5, NE:blue: n= 6). The dotted line shows the unity line. The results of a Spearman's Rank correlation are shown. (F) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a neuron in the (i) LHS or (ii) RHS, calculated across all detected neurons across all animals that met inclusion criteria. Results of a one-way ANOVA are shown, with alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$.

7.5.2.1 *Does lesion extent predict GFP labelling?*

The results presented so far suggest that lesions to CA1 do not prevent novel-exploration-dependent up-regulation of Arc-GFP activation in MEC layer V/VI. To test whether damage to the remainder of the hippocampus could influence the activity of Arc, I analysed the relationship between lesion extent throughout the dorsal, intermediate and ventral areas of CA1, CA2, CA3 and DG and GFP expression. There is no significant effect of lesion extent on soma-surround score (Fig.7.8). In fact, mice with almost the highest GFP antibody labelling have complete dorsal CA1 lesions and near-complete lesions throughout the whole of septo-temporal axis of CA1 (Fig.7.8(B)). However, there do appear to be trends in the effect of lesion extent to the whole of the CA1 region on Arc-GFP activation in novel-environment mice, with 5/7 mice showing higher expression in the hemisphere with less damage. There is no effect of lesion extent on Arc-GFP activation in home-cage animals.

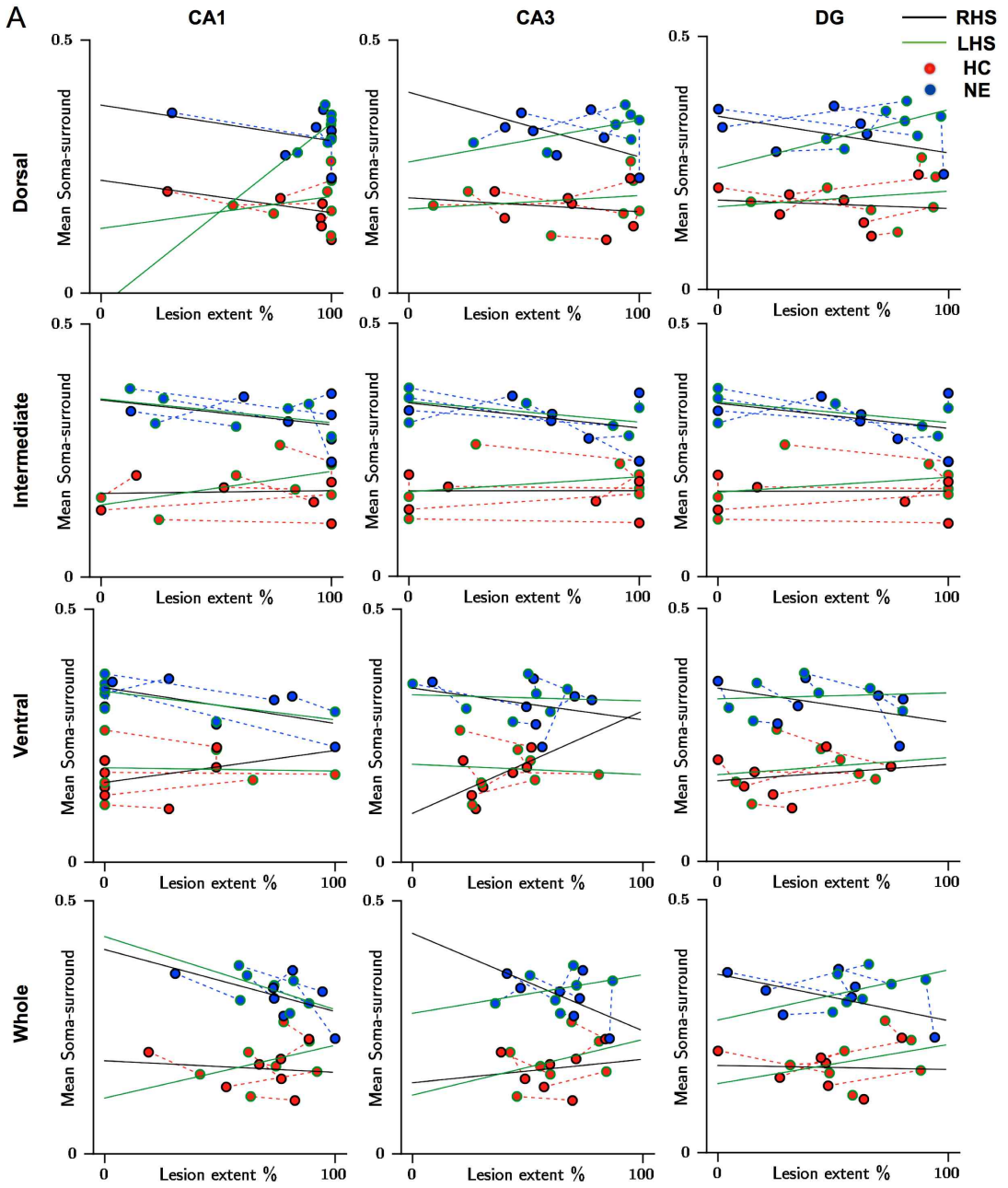
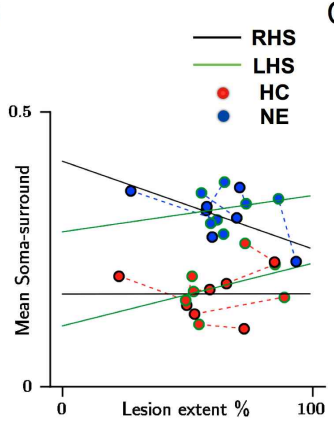


Figure 7.8: *Analysis of correspondence between hippocampal lesion extent and Arc expression in MEC layer V/VI.* The mean soma-surround score is plotted as a function of lesion extent for (A) GFP antibody labelling in the dorsal, intermediate and ventral extents of CA1, CA3 and the DG, and (B) the entire hippocampus. Data from both hemispheres are plotted and are joined by a dotted line for each mouse. (C) (below) Table showing correlations between lesion extent and Arc-GFP activation in each hemisphere for mice in each group. Red rows indicate home-cage animals.

B



C

Region	LHS Gradient	LHS r^2	LHS p	RHS Gradient	RHS r^2	RHS p
CA1d	-0.00064	0.15	0.390	0.00063	0.05	0.630
CA1d	-0.0007	0.14	0.410	0.0036	0.33	0.180
CA2d	0.00035	0.033	0.700	0.00038	0.1	0.490
CA2d	-0.00028	0.027	0.730	0.0009	0.54	0.060
CA3d	-0.00027	0.025	0.730	0.00027	0.043	0.660
CA3d	-0.0013	0.32	0.180	0.00083	0.43	0.110
DGd	-0.00017	0.013	0.810	0.0003	0.034	0.690
DGd	-0.00072	0.3	0.200	0.0012	0.37	0.150
CA1i	5.40E-05	0.0031	0.910	0.00066	0.28	0.220
CA1i	-0.00049	0.11	0.460	-0.00047	0.24	0.270
CA2i	0.0002	0.032	0.700	0.00064	0.33	0.180
CA2i	-0.00024	0.044	0.650	-0.00028	0.1	0.490
CA3i	0.00021	0.022	0.750	0.00089	0.28	0.230
CA3i	-0.00051	0.091	0.510	-0.00035	0.079	0.540
DGi	1.60E-06	3.30E-06	1.000	0.00029	0.087	0.520
DGi	-0.0005	0.13	0.420	-0.0004	0.32	0.180
CA1v	0.00063	0.12	0.450	-6.30E-05	0.0029	0.910
CA1v	-0.0007	0.34	0.170	-0.00056	0.42	0.110
CA2v	0.00049	0.32	0.190	-5.90E-05	0.0036	0.900
CA2v	-0.00058	0.23	0.280	-0.00058	0.47	0.088
CA3v	0.002	0.34	0.170	-0.0002	0.0076	0.850
CA3v	-0.00062	0.084	0.530	-0.00012	0.007	0.860
DGv	0.00032	0.037	0.680	0.00034	0.027	0.720
DGv	-0.00067	0.18	0.340	0.00012	0.0094	0.840
CA1w	-0.00023	0.017	0.780	0.001	0.14	0.400
CA1w	-0.0012	0.33	0.170	-0.0014	0.28	0.220
CA2w	0.001	0.19	0.330	0.00067	0.2	0.320
CA2w	-0.0006	0.099	0.490	0.00097	0.16	0.380
CA3w	0.00046	0.027	0.720	0.0011	0.14	0.410
CA3w	-0.0019	0.41	0.120	0.00076	0.13	0.420
DGw	-7.90E-05	0.0023	0.920	0.00077	0.11	0.480
DGw	-0.00092	0.33	0.170	0.00099	0.18	0.350
Hip	2.10E-06	1.00E-06	1.000	0.0011	0.16	0.370
Hip	-0.0016	0.45	0.100	0.00066	0.041	0.660

7.5.3 *Novel exploration causes CA1-independent up-regulation in Arc-GFP activation in visual cortex*

I have shown previously that novel exploration causes a robust increase in Arc expression in MEC layer V/VI and a detectable increase in the deep layer of visual cortex, layer VIa. Here I question whether this up-regulation is affected by lesions to CA1.

The number of detectable cells in visual layer VI increases following novel exploration (Fig.7.9(A-B)). This is significant for cells with a 'detectable' level of GFP fluorescence (Fig.7.9(C): t-test: $p = 0.016$), and showing moderate expression (t-test: $p = 0.025$), but not high expression (t test: $p = 0.098$). Detected cells appear to be highly similar across the two groups, which suggests that Arc-GFP activation occurs in a subset of cells at baseline that increases in size with novel exploration (Fig.7.9(D)).

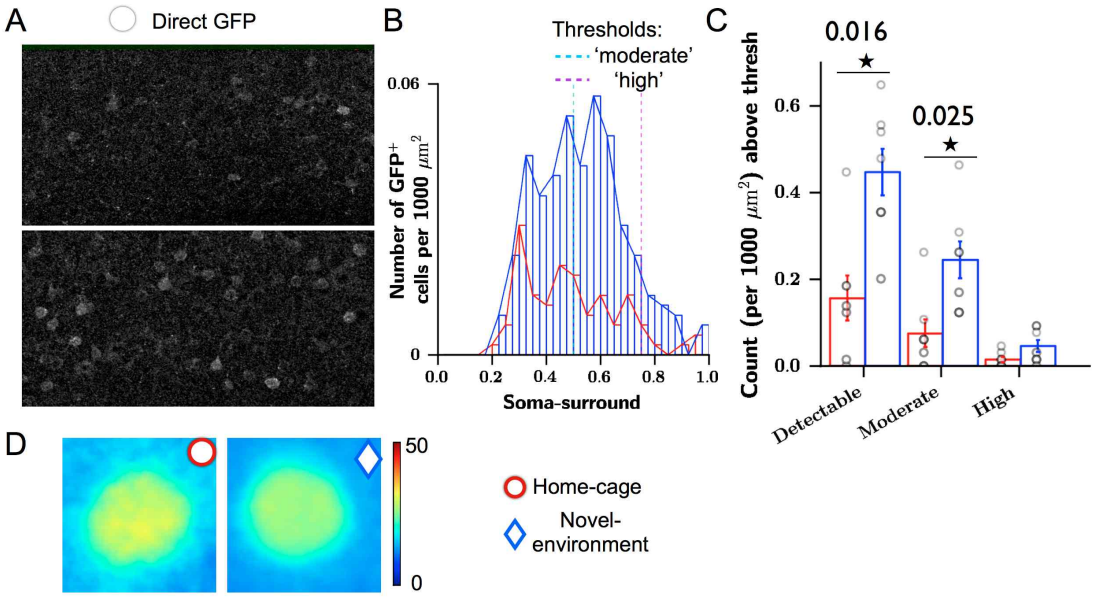


Figure 7.9: *Analysis of directly imaged GFP fluorescence in visual layer VI in home-cage and novel-environment animals following CA1 lesions* (A) Examples of confocal 20x maximum projection images from a home-cage (red circle) mouse and a novel-environment (blue triangle) mouse. (B) Histograms show the distributions of soma-surround score for cells detected using direct GFP fluorescence. Vertical lines mark the thresholds for the data plotted in (C). (C) The number of neurons exceeding thresholds marked in (B) is shown for home-cage and novel-environment animals. (D) Images show the average 60x60 pixel region of fluorescence surrounding a detected GFP^+ cell, calculated across all detected cells across each group of animals that met inclusion criteria. All data taken from the left hemisphere. Results of a one-way ANOVA are shown, with alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$. n = (hc: 6, ne: 6)

GFP antibody labelling measures also show changes following novel exploration in that a higher number of cells appear to be labelled and the distribution of soma-surround scores shifts to higher values (Fig.7.10(A-B)). Across the group there is an increase in the mean soma-surround score (Fig.7.9(C)(a): t-test: $p = 0.044$), and the variability in expression between neurons (Fig.7.9(C) (b): t-test: $p = 0.011$). It follows that there is an increase in the number of neurons with moderate and high levels of GFP labelling (Fig.7.9(D): t test: $p = 0.016$, $p = 0.027$). The effect of novel exploration on an average cell taken from any mouse in each group can be seen in Fig.7.9(E).

In summary, in visual cortex layer VI novel exploration continues to increase the number of cells that express levels of Arc within certain high ranges.

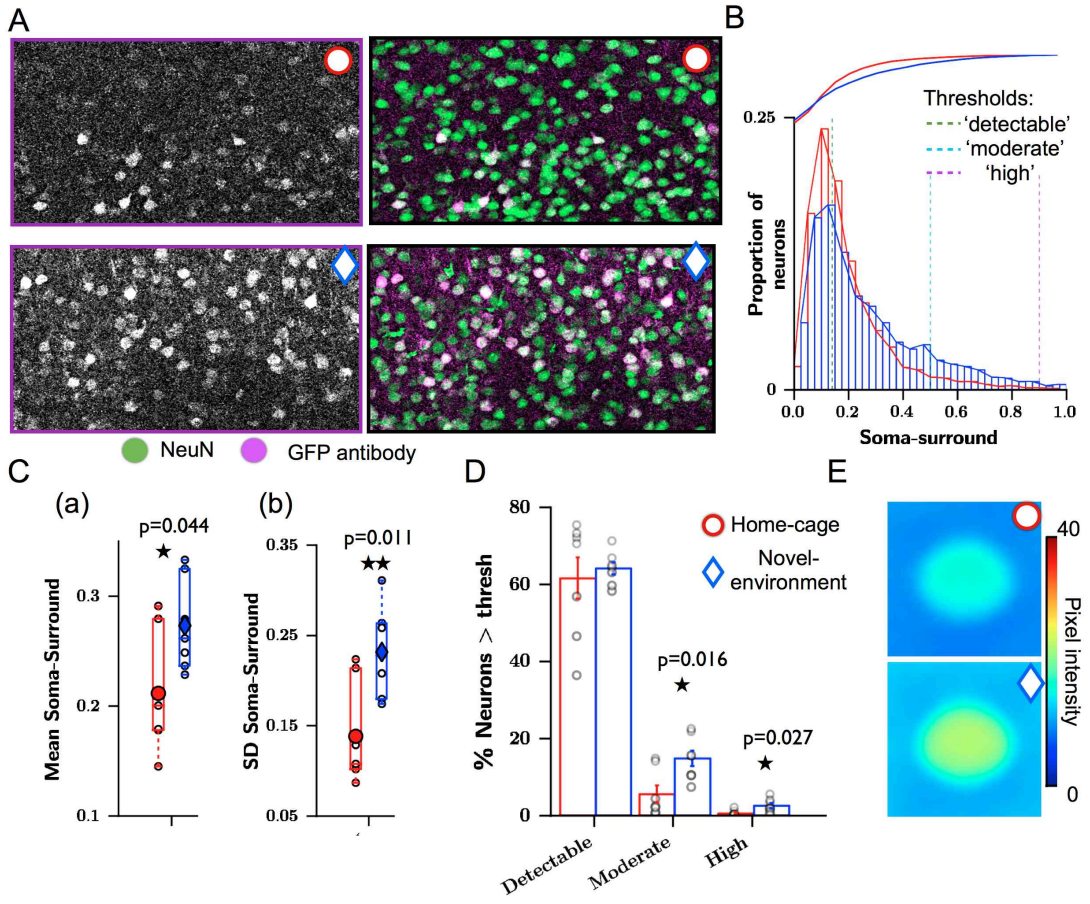


Figure 7.10: *Analysis of neuronal GFP antibody labelling in visual layer VI in home-cage and novel-environment animals following CA1 lesions* (A) Examples of confocal 20x maximum projection images from a home-cage (red circle) and novel-environment (blue triangle). GFP antibody and NeuN labelling are shown outlined in magenta and green, respectively. The accompanying image shows the GFP/NeuN overlay. (B) Histograms show the distributions of soma-surround score for all neurons detected using NeuN labelling. Vertical lines mark the thresholds for the data plotted in (D). (C) Boxplots show (a) The mean GFP labelling soma-surround score for all neurons, and (b) the standard deviation in soma-surround score. Filled shapes indicate the group mean. (D) The proportion of neurons exceeding GFP labelling thresholds marked in (B) is shown for home-cage (red) and novel-environment (blue) animals. (E) Images show the average 60x60 pixel region of GFP antibody labelling surrounding a neuron, calculated across all detected neurons in animals that met inclusion criteria for each group. Results of a one-way ANOVA are shown, with alpha levels of: * $p < 0.05$, **$p < 0.01$, *** $p < 0.001$ n = (hc: 6, ne: 6).

7.6 DISCUSSION

In this chapter I have shown that the increase in the expression of the activity-regulated gene, Arc, in dorsal MEC layer V/VI following novel exploration does not depend on input from dorsal CA1, from which it receives a strong projection. Arc-GFP activation shows a novelty-dependent increase in the MEC layer V/VI (Figs. 7.6,7.7) and in layer VI of visual cortex (Figs. 7.9,7.10). These changes are particularly concentrated at the upper end of the distribution, affecting the proportion of moderate to high-expressing cells, but not changing Arc expression uniformly across all neurons. In visual cortex the effect of novel exploration on Arc-GFP activation is less robust, which is consistent with the effects on Arc expression in non-lesioned animals (see previous chapter).

7.6.1 *What drives increases in Arc expression in MEC layer V/VI?*

7.6.1.1 *From an anatomical and physiological perspective*

The result that novel exploration continues to increase Arc expression in MEC layer V/VI following dorsal CA1 ablation is initially surprising because the traditional view is that CA1 provides a major input to MEC layer V/VI (Swanson and Cowan, 1977), which thens relays information to neocortex (Insausti et al., 1997). However, the data presented in the previous chapter that show that Arc-GFP activation does not increase in CA1 or in layer II of the MEC, but does increase in layer V/VI, suggests that some information processing in layer V/VI could be independent of other hippocampal-entorhinal processing. One possible explanation for the previous result was that although activity in the entorhinal-hippocampal circuit may drive Arc expression in layer V/VI neurons, the criteria for an increase in Arc expression may not have been met in this experiment. For example, although Arc is activity-regulated, its activation is delayed relative to other immediate early genes and may depend on more factors, such as NMDA signalling (Lyford et al., 1995; Bramham et al., 2010). However, the data presented here suggest an alternative explanation. Up-regulation of Arc in MEC deep layers does not depend on activity in CA1.

So which inputs to layer V/VI neurons stimulate the combined actions of the transcription factors that regulate Arc expression during novel exploration, and what are the behavioural requirements? Recent research considering the morphology and inputs to these cells suggests that there could be many anatomical candidates (Medinilla et al., 2013). Layer V pyramidal neurons possess a long apical dendrite that extends into the superficial layers of MEC (Hamam et al., 2002). These dendrites are spiny throughout (Hamam et al., 2002; Medinilla et al., 2013; Gasparini, 2011) and therefore have the potential to receives inputs that enter MEC across all layers. One of the more proximal input sources, aside from the hippocampal and subicular input, originates in the retrosplenial cortex, which may process head-direction information during navigation (Chen et al., 1994). The MEC also receives inputs from post, pre and parasubiculum, claustrum, and many neocortical regions (Canto et al., 2012; Burwell and Amaral, 1998). These inputs could interact to provide a consolidated input to MEC neurons that stimulates the engagement of plasticity mechanisms.

Plasticity in MEC has not been extensively studied and its function is not known, but the ability of grid fields to remap in novel environments (Fyhn et al., 2007) suggests that plasticity may be an important feature of the circuit dynamics. CA1-MEC layer V synapses show capability for short-term and long-term plasticity and a frequency-dependence that will switch potentiation to depression (Craig and Commins, 2006, 2007), but it is not clear whether this plasticity, studied *in vitro*, could stimulate Arc. Although the inputs to MEC layer V neurons have not been well characterised, their electrophysiological characteristics, specifically their response to depolarising inputs, have led to previous proposals that they may be important for working memory (Egorov et al., 2002; Fransén et al., 2006), which could explain a requirement for Arc.

7.6.1.2 *From a behavioural perspective*

From a behavioural perspective, which regions are active during novel exploration, provide inputs to MEC deep layers and could engage cellular mechanisms that up-regulate Arc expression without hippocampal input? Experiments have shown that the formation of memories that require the learning of associations between object and place depend on the hippocampus (Langston and Wood, 2010). Hippocampal lesions also disrupt spatial navigation in the water maze task, preventing rats from remembering the location of a hidden platform (Morris et al., 1982; Moser et al., 1993), and they disrupt spatial reference and working memory in the radial maze in mice (Murray et al., 2011). It is therefore unlikely that Arc expression in MEC deep layers is uniquely related to these types of learning.

Novel exploration of a spatial environment incorporates a number of psychological and behavioural aspects that could increase Arc. In this experiment the animals were completely naive to the novel environment in which they were placed. They had never been in the room, had never experienced such colourful and interesting objects and had never been able to run over such a large area of space. Such an experience probably stimulates not only the encoding of novel associations between objects and their spatial environment, but also the signalling of potential stress and the recruitment of attentional and working memory mechanisms for processing the available information. Working memory tasks that involve long delays have been shown to depend on interactions between the hippocampus and prefrontal cortex, but memory over short delays can be supported by individual regions (Churchwell and Kesner, 2011). Given that MEC layer V has been implicated in working memory previously (Egorov et al., 2002), it is plausible that Arc expression either supports or is caused by related mechanisms. The deep layers receive inputs from a wide range of areas so are ideally positioned to integrate this information during an experience such as novel exploration.

7.6.2 *Hippocampus-independent visual cortex up-regulation*

I initially predicted that CA1 would play a more significant role in novel-exploration dependent Arc up-regulation in the deep layers of MEC than in visual cortex. Although visual cortex receives inputs from the hippocampus via MEC layer V (Insausti et al., 1997), it also receives a more direct input. Previous work has shown that in posterior parietal cortex and granular insular cortex, hippocampal lesions (predominantly dorsal) do not change the pro-

portion of neurons expressing Arc following traversal of multiple tracks in different rooms (Takehara-Nishiuchi et al., 2012). The results here are therefore partially consistent with this prediction in that I find, particularly using direct GFP fluorescence (Fig. 7.9) that Arc-GFP activation in visual layer VI continues to increase following novel exploration. In contrast, hippocampal lesions or inactivation do affect immediate early gene expression in another region of cortex, the retrosplenial cortex (Kubik et al., 2012; Albasser et al., 2007). Lesions halve the number of superficial cells expressing immediate early genes following behavioural tasks (Albasser et al., 2007), while TTX infusion completely disrupts the activity-dependent increase in Arc expression (Kubik et al., 2012). This region may therefore depend on different inputs from the entorhinal-hippocampal system to the visual cortex.

7.6.3 *Potential limitations*

Lesions using an excitotoxic agent are just one method of disrupting a system to test whether it is necessary for a particular function. The advantage of such lesions are that their extent can be directly analysed and they can be administered over reasonably large areas of space. However, lesions must be performed in advance of testing, in this case 1 week, which could provide time for the system to adapt and for the importance of other inputs to MEC deep layer cells to change. It is also difficult to lesion regions with high specificity because of slight variability in the positioning of structures between mice.

To complement the results from this lesion study and show both that the system has not adapted over time and that the lesion has successfully disrupted it, I could perform several experiments. First, instead of lesions I could use temporary inactivation of the hippocampal region by applying, via a pre-implanted cannula, an agent such as muscimol, a GABA agonist, while the mouse explores a novel environment. These results would confirm that the result is not due to a long-term adaptation in the system, although assessing and confirming the extent of inactivation is a challenge. Second, I could test the hippocampus-lesioned mice on a hippocampus-dependent task such as an object-place-context task (Langston and Wood, 2010). If lesioned mice were unable to perform this task, the results would show more convincingly that the exploration-induced up-regulation of Arc in MEC deep layers does not, in the typical hippocampal system, depend on hippocampal processes that are necessary for spatial associative memory.

In this experiment I chose not to use a sham control group that received a PBS injection instead of ibotenic acid because pilot experiments indicated that the measurement of interest, novel-exploration induced Arc up-regulation, persists in animals with dorsal CA1 lesions. Sham-operated animals provide a control for effects of the surgical procedure, but cannot provide information about a lack of effect. By using lesioned home-cage animals as a control, I could confirm in this experiment that lesions do not prevent novel-exploration induced increases in Arc and that other inputs likely contribute to this. A more informative control than a sham would test the success of the lesions, as discussed above. However, sham control animals would have enabled me to address a secondary question of quantifying precisely whether dorsal CA1 lesions influence Arc expression in the MEC deep layers, even if they do not prevent it. Because I used animals from a different breeding regime and I did not match

animals across lesioned and non-lesioned groups through such a control, I cannot clearly establish this. Lesions both reduce afferent input to MEC deep layers, even if this is not an important factor in regulating Arc promoter activity, and also change an animal's behavioural experience of the novel exploration. This is particularly true given that in some cases a large area of the hippocampus was lesioned. Ventral CA1 lesions have been shown to reduce anxiety (Kjelstrup et al., 2002), which could affect the extent to which animals explore, although I observed no evidence of this. Previous work lesioning the hippocampus has shown that spatial learning is impaired following lesions of just 20% of the dorsal hippocampus and that impairment correlates with size of lesion (Moser et al., 1993). To investigate whether exploration-induced Arc-GFP activation could be influenced by lesions, I correlated it with lesion extent throughout different regions of the hippocampus. For most hippocampal regions there is no evidence of a correlation but there is some, non-significant evidence in the data that Arc-GFP activation is affected in animals with unequal lesions of ventral or the entire of CA1 in the two hemispheres. It is therefore possible that ventral CA1 processing influences Arc expression in dorsal deep MEC layer networks. This experiment certainly cannot rule this possibility out, although such lesions do not appear to prevent exploration-induced Arc up-regulation. To test whether lesions extending over ventral, dorsal and the entire of CA1 have different effects on Arc expression, it would be necessary to explicitly lesion these regions separately.

7.6.4 *Open Questions*

Despite showing the important result that novelty-dependent Arc up-regulation in MEC layer V/VI does not depend on CA1, this experiment leaves several important questions unanswered. I have focussed on septal (dorsal) CA1, first because septal hippocampal exchanges reciprocal connections with dorsal MEC layer V, whereas temporal CA1 is interconnected with ventral MEC (van Groen et al., 2003; Cenquizca and Swanson, 2007), and second because CA1 provides the major output from the hippocampus. Although early reports suggested that CA3 and dentate gyrus send projections to subiculum that bypass CA1 (Swanson and Cowan, 1977) and that could therefore reach layer V/VI independently of CA1, more recent work has shown that CA3 does not connect to the subiculum (Witter & Amaral, as cited in (Witter, 2007)). The early results may have been due to spread of tracers into neighbouring regions. Because I have not explicitly tested the effect of CA3 or dentate gyrus lesions, their role in Arc expression in MEC deep layers is unclear but unlikely to be influential given the lack of correlation between lesion extent and Arc-GFP activation (Fig. 7.8). In addition, I observed examples of higher levels of GFP in a novel-environment animal than a matched baseline animal despite near-complete hippocampal lesions.

In the analysis used here I have briefly considered differences between the left and right hemisphere of MEC layer V/VI. In the human hippocampus, there is evidence that spatial navigation may be more lateralised to the right than left hemisphere (Burgess et al., 2002) and in rats the number of genes that show changes in expression following the learning of a spatial reference memory task is higher in the right than left hemisphere (Klur et al., 2009). I found no consistent differences between Arc expression in the left and right hemispheres

even in animals with lesions with very similar extent (Figs 7.8,7.7). However, since lesions to the two hemispheres were not selectively controlled, it is difficult to conclude that there are no differences in the dependence of MEC layer V/VI Arc expression across the two hemispheres. It is known that left and right MEC have contralateral projections and that grid cell organisation shows similar modularity properties across both hemispheres (Stensola et al., 2012) so the similar results across the hemispheres could reflect the fact they are processing coordinated information.

Inactivation of the hippocampus disrupts grid field activity in MEC layer II (Bonnevie et al., 2013), via pathways that could either include the MEC deep layers, or that could lead directly from the CA2 region to MEC (Rowland et al., 2011). I did not consider the effect of hippocampal lesions on Arc-GFP activation in MEC layer II because it is low and is not increased by novel exploration. However, lesions could affect novel-exploration induced Arc protein expression in layer II. If this were the case, it might indicate that the latter pathway is more likely.

7.6.5 *Conclusions*

This research has robustly demonstrated that Arc-GFP activation in layers V/VI of MEC increases following exposure to a novel environment in the absence of the dominant inputs from dorsal CA1. It therefore does not rely on this input. Many of the tools used for understanding MEC circuitry *in vivo* have very low sampling rates because the positioning of MEC makes *in vivo* imaging difficult. Tetrodes provide information concerning a large number of cells, but it is difficult to simultaneously achieve this throughout MEC layers and the dorsoventral axis. The measurement of activity-dependent gene expression therefore provides a complementary insight into the plasticity-related activity of cells across a wide area. This work indicates that important plastic changes may be occurring onto layer V/VI neurons during navigation that are independent of the hippocampus. The next challenge is to establish which inputs provide the stimulus for Arc expression, which cell types are receiving them, and whether they are functionally important.

8

GENERAL DISCUSSION

This thesis has explored the molecular profile of the medial entorhinal cortex (MEC) and sought to map patterns in gene expression in MEC, at baseline and after exploration of a novel environment, to anatomy, physiology, function and pathology. The MEC is a fascinating structure, connecting extensively and directly with a large number of regions throughout the brain, including the hippocampus. Its connections with the hippocampus are important for long-term spatial memory (Remondes and Schuman, 2004), as well as other forms of memory (reviewed in Eichenbaum and Lipton, 2008), and it contains cells that provide a grid-like spatial readout of an animal's environment (Hafting et al., 2005). One of the biggest challenges for understanding how the MEC contributes to spatial learning and navigation is specifically targeting, manipulating and disrupting aspects of the circuit to test hypotheses. The MEC is also of neuroscientific interest because it shows pathology in a number of diseases, most notably Alzheimer's disease (Braak and Braak, 1991; Gómez-Isla et al., 1996), but it remains unclear what confers its vulnerability. Alzheimer's disease is the major cause of dementia and presents many social and economic challenges. It is therefore of paramount importance that we develop an understanding of what triggers neurodegeneration in subregions of MEC and how this could be controlled.

The MEC circuit itself has a laminar structure, but a considerable number of cell types are intermingled throughout layers and our understanding of their individual contributions are limited. Given the distinctive molecular profiles of certain neocortical, hippocampal (Sugino et al., 2006), retinal (Siegert et al., 2012) and striatal cells (Doyle et al., 2008), and of neocortical layers (Belgard et al., 2011), it seems likely that MEC layers, and even cell types, may also be distinguishable on the basis of gene expression. Mapping gene expression to function has two key advantages in this way: (1) isolating genes that are specific to certain layers or cell types provides a tool for selective targeting, and (2) genes encode proteins that perform functions in the cell that may help to understand cellular and circuit function. It is possible that gene expression could also provide insight into the vulnerability of MEC, layer II in particular, to Alzheimer's disease pathology.

To my knowledge, I have provided the first extensive and detailed characterisation of the localisation of RNA to highly specific MEC regions, using a purely *in silico* analysis of ISH data from the Allen Brain Atlas. I have shown that a considerable proportion of genes are expressed non-uniformly across MEC layers, with over 100 showing highly selective expression to a single layer. Specific genes are most often restricted to layer II of MEC, where the largest proportion of grid cells are found (Sargolini et al., 2006). I have also shown that genes that are expressed in MEC are not always distributed uniformly throughout the dorso-ventral extent, but a considerable proportion are dorso-ventrally organised in the superficial layers, such that dorsal cells express higher levels than ventral cells. The dorso-ventral axis of MEC has received considerable attention since it was shown that the spatial resolution of grid fields

is topographically organised along it (Hafting et al., 2005; Brun et al., 2008b; Stensola et al., 2012). The synaptic integrative properties of stellate cells also follow a gradient (Garden et al., 2008), so dorso-ventrally organised genes could both provide insight into what makes cells different, and provide tools for manipulating the gradient.

One group of genes with a particularly close mapping to function (Leslie and Nedivi, 2011) that is overrepresented amongst dorso-ventral genes is the activity-regulated immediate early genes. I therefore asked whether one such gene, the activity-regulated cytoskeleton-associated gene, Arc, might provide insight into MEC circuit organisation in behaving animals. To address this, I investigated changes in Arc expression following exposure to a novel spatial environment and showed using an Arc-GFP reporter mouse line, in which GFP is under the control of the endogenous Arc promoter (Wang et al., 2006), that novel exploration does not significantly change Arc-GFP activation in layer II, but does modestly increase Arc protein expression. By contrast, Arc-GFP activation and protein expression are both significantly increased in the MEC deep layers, an effect which is still evident when the CA1 input to the deep layers is removed using a targeted lesion. These results provide the first indication that deep layer neurons perform a potentially important role, *in vivo*, in the encoding of novel spatial environments.

I have therefore shown that both baseline and activity-dependent gene expression provide insight into function in MEC circuits. In the following few sections I have identified and outlined the features of this thesis that I believe contribute most to neuroscience research and, in particular, to our understanding of the MEC: the use of in silico methods to characterise the transcriptomic profile of regions, the functionally relevant enrichment of genes in MEC layer II, the contribution of dorso-ventrally organised genes to function, laminar differences in the expression of Arc during novel exploration in MEC, and the role of the MEC deep layers.

The strategy for determining laminar differences in gene expression using high-throughput ISH tools

In this thesis I have presented a strategy for extracting gene expression data from freely available ISH images in an online database and using it both to identify genes with highly specific patterns of expression and to present global differences in the types of genes enriched in layers of MEC. This strategy is both widely applicable in the sense that it could be applied to multiple different brain regions, and highly flexible in the types of questions that it can be used to ask. For example, researchers could use this approach to finding layer-specific genes in any region of neocortex or subcortical structures. As demonstrated in this work, it is possible to extract gene expression patterns from MEC, visual cortex and somatosensory cortex layers, but other regions could also be investigated. Such data can be used to compare brain regions and identify those that show particular similarities to one another. I showed here that MEC deep layers exhibit gene expression that is most consistent with neocortical layer VIa, whereas gene expression in superficial MEC is more similar to superficial neocortex. It is also possible to identify and measure gradients throughout different brain regions. In this analysis I have only considered linear gradients, but exponential gradients and step functions could also be tested for. This strategy could even be expanded to investigating differences between gene expression at different developmental stages, and in different organisms. Im-

portantly, although it has limitations, the advantage of this strategy over some existing ABA tools is that it is possible to perform automated error checks that indicate the likelihood that a particular result is valid. These types of approaches can be combined with physiology and anatomical analysis to improve our understanding of neural circuits and what distinguishes their components.

Gene expression in MEC layer II

Neurons in layer II of MEC have received considerable attention previously; a large proportion are spatially modulated (Sargolini et al., 2006; Zhang et al., 2013), they exhibit a gradient in morphological and electrophysiological properties (Hafting et al., 2005; Garden et al., 2008) and a subset have the potential to form synapses onto newly born granule cells (Nakashiba et al., 2012). They are also susceptible to pathology in Alzheimer's disease (Braak and Braak, 1991) and other neurodegenerative diseases (Braak et al., 2000), schizophrenia (Arnold, 2000), and, to a lesser degree, epilepsy (Du et al., 1993).

I have now shown that layer II neurons are also the most likely to show high, enriched gene expression compared with other MEC layers, and that many of the over 1000 enriched genes are dorso-ventrally organised. Layer II-enriched genes do not always follow the same pattern of expression but at least two different patterns can be reliably identified that correspond to the expression patterns of the calcium binding protein, calbindin, and reelin. Layer-II enriched genes are overwhelmingly enriched for mitochondrial functions, similar to layer V cells of the somatosensory (SS) cortex, which are large and send long-range projections outside of the skull (Belgard et al., 2011). Layer II neurons are also large, but what drives a high energy requirement is not immediately clear. In SS cortex, mitochondrial genes are expressed at the highest levels (Belgard et al., 2011), but it is not easy to quantify this for MEC with ISH data.

Layer-II enriched genes are also enriched in the post-synaptic LTP pathway, which suggests that plasticity is an important feature in layer II cells. It could be relevant to their ability to process inputs from diverse regions and the ability of grid cells to remap in novel environments (Barry et al., 2007; Fyhn et al., 2007). Intriguingly, I found that genes that are specific to layer II, within MEC, are enriched for axonogenesis-related molecules, which may be critical to their ability to develop new synapses with granule cells. The identity of these genes is therefore useful for those wanting to understand the importance of this connection. The final observation regarding baseline gene expression in layer II is that genes encoding proteins that are perturbed or targeted in neurodegenerative pathways are overwhelmingly overrepresented. Given that Alzheimer's disease pathology is evident early on in the MEC layer II (Braak and Braak, 1991), this combination of gene enrichment, if disrupted, could be the cause of vulnerability. It will be important to investigate how the expression of these genes, both RNA and protein, changes with ageing and to potentially develop mouse models in which protein pathways that are encoded by genes enriched in layer II are disrupted.

The contribution of dorso-ventrally organised genes to function

The dorso-ventral organisation of anatomical, physiological, and molecular properties of MEC is an intriguing feature because of the influential roles played by molecular gradients in the spinal cord and brain during development. I found that in the MEC the organisation is

overwhelmingly biased towards high dorsal expression. While cell density or size may factor in this, the large proportion of genes with no gradient in expression throughout suggests that this cannot fully explain the results. Genes related to ion channel activity are highly likely to be enriched in dorsal cells, whereas post-synaptic density genes appear to distinguish genes with different patterns of expression. Very few genes show the opposing ventro-dorsal gradient, particularly in the superficial layers. However, such genes do exist, as do genes that are expressed at selective points along the dorso-ventral extent. Cells that likely contribute to spatial navigation and learning are organised in a modular manner along this axis (Stensola et al., 2012), an arrangement that could, in theory, be genetically encoded. However, the rarity of these genes and their variable expression pattern suggests that at the level of RNA, at least, this may not be the case. Nevertheless it is possible that measuring the expression level of multiple genes, as well as protein, simultaneously in the same mouse could reveal the modular code.

The MEC circuit during novel exploration

This is the first experiment, to my knowledge, in which laminar differences in the somatic expression of an activity-dependent gene have been reported in MEC. Previous research has demonstrated that novel exploration and activity increase the level of Arc RNA in ventral MEC deep layer tissue (Gusev et al., 2005), but it was not quantified at a laminar or neuronal level. I have shown that when a mouse is placed in a novel environment, Arc is active at the highest levels in dorsal layer V/VI neurons. Exploration also considerably increases the protein expression of Arc in cell bodies. This suggests that Arc, a protein required for various forms of plasticity (summarised in (Bramham et al., 2010)), is required at synapses onto these neurons. By contrast, novel exploration does not significantly increase Arc-GFP activation in layer II in an Arc-GFP reporter line, despite evidence that layer II cells are active when an animal explores a novel environment (Burgalossi et al., 2011). I did, however, find evidence that Arc protein expression increases. It would be interesting to discover whether the physiological Arc-dependent effect on layer II and V/VI cells is similar following this experience, or whether different signalling pathways are affected. I did not have sufficient data to test whether there is a gradient in Arc protein expression, or whether this is modulated by exposure to a novel environment, but it would be interesting to determine whether different types of novel exploration, or sizes of novel environment differentially affect any gradient in the promoter activity and translation of Arc.

The contribution of the MEC deep layers

The MEC deep layers are positioned as a gateway from the hippocampus to the neocortex and other cells within MEC (Swanson and Cowan, 1977; Tamamaki and Nojyo, 1995; Canto et al., 2008). Although information can also be routed directly from the hippocampus to these regions, it is likely that this pathway serves an important purpose in modulating information and in providing feedback. The MEC deep layers may also process information independently of CA1 inputs, as suggested by the novel exploration-induced increase in Arc expression in layer V/VI cells in the absence of this input. However, many different cell types exist throughout the deep layers, some of which are intermingled (summarised in (Canto et al.,

2008)), so it is difficult to isolate particular types and establish function, or to establish which cell types are up-regulating Arc expression. I showed in Chapter 2 that a considerable number of genes are uniquely expressed in at least one of the deep layers, which also receive direct input from retrosplenial areas, thalamic areas and other regions of MEC. Since a subset of these genes appear to segregate the region, they could provide a more systematic means of identifying neurons that also show particular morphological and connectivity features so that they can be targeted more effectively.

FUTURE WORK

The element of this work that perhaps carries the most potential is the identification of selective genes that are of functional interest. There is considerable evidence that data in the Allen Brain Atlas is accurate and reproducible. Nevertheless, a critical step if one is to investigate whether the genes identified here are actually specifically expressed in particular MEC layers, or even exclusively in a particular neural subtype, is to confirm this using *in situ hybridisation*, or RNA-seq, and to test whether the encoded protein is also specific using immunohistochemistry. I demonstrated using an Arc-GFP knockin line, in which GFP is under the control of the endogenous Arc promoter, that in some MEC cells novel exploration may differentially affect the activation and protein expression of Arc. I also provided evidence, through the use of Arc knockout animals, that loss of Arc protein could actually lead to increases in the activity of its promoter, but this requires further testing. Given the usage of RNA expression, both at baseline and following activity, as an indicator of function, it will be important to establish exactly what the relationship between RNA expression and protein is for genes that appear to be specifically expressed.

FINAL CONCLUSIONS

The MEC layers are functionally and molecularly distinct. I have identified genes with highly selective expression and provided insights into how gene expression might contribute to laminar function. Returning to the first chapter and the relationship between understanding gene expression and connectivity, although tracers have indicated a huge array of possible connections in the MEC circuit, very few are known to be functional and even fewer can be selectively targeted. The identification of specific molecules could help to establish the identity of neurons that actually form synaptic connections and how these neurons function in the baseline and active, behaviourally modulated circuit.

APPENDIX: EXPERIMENTAL PROTOCOLS

A.1 GENOTYPING

Genotyping protocol for Arc-GFP mice

Reagents:

1. GoTaq Green 2x (contains 3mM MgCl2)
2. 0.69M Betaine (for amplification)
3. Primers

Make up 12µl mastermix (add 1 to sample number) - make up 2 reactions (1 WT, 1 mut) for each sample:

1. 6µl GoTaq Green - first thaw and vortex
2. 0.1µl WT/Mutant Fwd - pulse spin
3. 0.1µl WT/Mutant Reverse - pulse spin
4. 0.075µl Beta-globin Fwd
5. 0.075µl Beta-globin Reverse
6. 1.66µl Betaine
7. 3.79µl dH20

PCR mix:

1. Add 11.8µl mastermix to each PCR tube
2. Add 0.2µl DNA (extracted using DNAreleasy)
3. Pulse spin

Cycling

Step #	Temp °C	Time	Note
1	94	3 min	-
2	94	30 sec	-
3	64	1 min	-
4	72	1 min	repeat steps 2-4 for 35 cycles
5	72	2 min	-
6	10	-	hold

Notes:

Always use filtered tips

Run 5µl on 1.5% agarose gel

A.2 GFP IMMUNOFLUORESCENCE

NeuN/GFP staining protocol

For free-floating - use 24 well plate, 500ul per well for up to 4 sections, 250ul for 1-2 sections per well

1. Wash sections for 15 mins in PBST (I use 0.4% Tween20) at RT on shaker
2. Block in 10% Normal Goat Serum (Jackson Immuno) in PBST for 1 hour at RT on shaker
3. Incubate in 1:500 Millipore Mouse anti-NeuN in PBST 1% NGS, 1:2000 rabbit anti-GFP antibody for 3 hours at room temp. (RT) on shaker
4. Wash 3 x 5-10 mins in PBST at RT on shaker.
5. Incubate 2 hours in Alexa Fluor Goat anti-mouse 647 (1:400)/ Alexa Fluor Goat anti-rabbit 546 (1:400) in PBST, 1% NGS RT on shaker
6. Wash at least 3 x 20 mins in PBS at RT on shaker
7. Mount in Vectashield with coverslip of 1.5 thickness

A.3 ARC IMMUNOFLUORESCENCE

Protocols: Anti-Arc C7 GaM IF, free floating

Use 24-well plates
Use 500µl per well (4 sections)

Day 1:

1. Wash sections for 15 mins in PBST (0.1% Tween20)
2. Antigen retrieval
    a. Heat 10mM sodium citrate, 0.1% Tween, pH6 to 80 (check this temperature) in 2ml eppendorf in heat block.
    b. Add sections
    c. 30 min incubation
    d. Remove from heat block and leave to cool in citrate buffer
3. Block in 10% Normal Goat Serum (Jackson Immuno) for 1 hour at RT (room temp.) on shaker
4. Incubate sections in primary mouse Arc C7 (Santa Cruz) overnight (1:100) in PBST, 1% NGS, at 4° on shaker.

Day 2:

5. Wash 3x5-10 minutes with PBS-Tween 20.
6. Incubate 1:100 in Alexa Fluor Goat anti-mouse 647 (Invitrogen) in PBST, 1% NGS for 2 hours at RT on shaker
7. Wash 2 x 15 mins in PBS
8. Incubate 30 mins in NeuroTrace 1:100 on shaker at RT
9. Wash 2 x 5 mins in PBS
10. Wash 1 hour in PBS
11. Mount in Vectashield Hard set using 1.5 thickness coverslips and image as soon as possible

A.4 NISSL STAINING

Protocols: Nissl staining for hippocampal lesions

Notes:
Make up ETOH solutions with distilled water
Make up Cresyl Violet: add 1.25 g cresyl violet acetate and 0.75 mL glacial acetic acid to 250 mL of warm dH2O, then cool and filter (store unused in fridge) (http://neuroendocrinology.blogspot.co.uk/2012/01/recipe-cresyl-violet-staining-for.html)
Use slide boxes for incubation

Step 1: Perfusion, fixation, cryoprotection, cutting
Step 2: Mount using Superfrost microscope slides - DO NOT use old ones, the sections will slip off/get damaged. Mount every 6 sections (each 40 microns), so roughly every 250 microns
Step 3: (in minutes)

95% ETOH	15
70% ETOH	1
50% ETOH	1
dH20	2
dH20	1
Cresyl violet	15 (changed from 10)
dH2O	1
50% ETOH	1
70% ETOH + 1% acetic acid	2
95%	1
100%	1
Xylene	5

Step 4: Air dry
Step 5: Mount in Eukitt - ideally in fume cupboard - and coverslip with long coverslips. Use plastic disposable pipettes to place mountant onto slides. Coverslip quickly because it hardens quickly
Step 6: Imaging

LIST OF FIGURES

BIBLIOGRAPHY

Ainge, J. A., van der Meer, M. A. A., Langston, R. F., and Wood, E. R. (2007). Exploring the role of context-dependent hippocampal activity in spatial alternation behavior. *Hippocampus*, 17(10):988–1002. (Cited on page 14.)

Alarcón, M., Abrahams, B. S., Stone, J. L., Duvall, J. A., Perederiy, J. V., Bomar, J. M., Sebat, J., Wigler, M., Martin, C. L., Ledbetter, D. H., Nelson, S. F., Cantor, R. M., and Geschwind, D. H. (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *American Journal of Human Genetics*, 82(1):150–159. (Cited on page 12.)

Albasser, M. M., Poirier, G. L., Warburton, E. C., and Aggleton, J. P. (2007). Hippocampal lesions halve immediate-early gene protein counts in retrosplenial cortex: distal dysfunctions in a spatial memory system. *The European Journal of Neuroscience*, 26(5):1254–1266. (Cited on page 288.)

Alberi, L., Liu, S., Wang, Y., Badie, R., Smith-Hicks, C., Wu, J., Pierfelice, T. J., Abazyan, B., Mattson, M. P., Kuhl, D., Pletnikov, M., Worley, P. F., and Gaiano, N. (2011). Activity-induced Notch signaling in neurons requires Arc/Arg3.1 and is essential for synaptic plasticity in hippocampal networks. *Neuron*, 69(3):437–444. (Cited on page 38.)

Alonso, A. and Klink, R. (1993). Differential electroresponsiveness of stellate and pyramidal-like cells of medial entorhinal cortex layer II. *Journal of Neurophysiology*, 70(1):128–143. (Cited on page 22.)

Amaral, D. G. and Witter, M. P. (1989). The three-dimensional organization of the hippocampal formation: a review of anatomical data. *Neuroscience*, 31(3):571–591. (Cited on page 20.)

Arnold, S. E. (2000). Cellular and molecular neuropathology of the parahippocampal region in schizophrenia. *Annals of the New York Academy of Sciences*, 911:275–292. (Cited on pages 30, 42, 154, and 293.)

Arnold, S. E., Hyman, B. T., Van Hoesen, G. W., and Damasio, A. R. (1991). Some cytoarchitectural abnormalities of the entorhinal cortex in schizophrenia. *Archives of General Psychiatry*, 48(7):625–632. (Cited on pages 13 and 30.)

Bachevalier, J. and Nemanic, S. (2007). Memory for spatial location and object-place associations are differently processed by the hippocampal formation, parahippocampal areas TH/TF and perirhinal cortex. *Hippocampus*, 18(1):64–80. (Cited on page 14.)

Barry, C., Hayman, R., Burgess, N., and Jeffery, K. J. (2007). Experience-dependent rescaling of entorhinal grids. *Nature Neuroscience*, 10(6):682–684. (Cited on pages 17, 108, 215, and 293.)

Bartesaghi, R. and Gessi, T. (2004). Parallel activation of field CA2 and dentate gyrus by synaptically elicited perforant path volleys. *Hippocampus*, 14(8):948–963. (Cited on page 15.)

Bauman, M. and Kemper, T. L. (1985). Histoanatomic observations of the brain in early infantile autism. *Neurology*, 35(6):866–874. (Cited on page 13.)

Beed, P., Bendels, M. H. K., Wiegand, H. F., Leibold, C., Johenning, F. W., and Schmitz, D. (2010). Analysis of excitatory microcircuitry in the medial entorhinal cortex reveals cell-type-specific differences. *Neuron*, 68(6):1059–1066. (Cited on page 22.)

Belgard, T. G., Marques, A. C., Oliver, P. L., Abaan, H. O., Sirey, T. M., Hoerder-Suabedissen, A., García-Moreno, F., Molnár, Z., Margulies, E. H., and Ponting, C. P. (2011). A transcriptomic atlas of mouse neocortical layers. *Neuron*, 71(4):605–616. (Cited on pages 7, 42, 49, 135, 136, 150, 154, 291, and 293.)

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B ...*. (Cited on pages 114 and 125.)

Bertram, L. and Tanzi, R. E. (2008). Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature Reviews Neuroscience*, 9(10):768–778. (Cited on pages 10, 11, 42, and 153.)

Bird, C. M. and Burgess, N. (2008). The hippocampus and memory: insights from spatial processing. *Nature Reviews Neuroscience*, 9(3):182–194. (Cited on page 14.)

Bloomer, W. A. C., VanDongen, H. M. A., and VanDongen, A. M. J. (2008). Arc/Arg3.1 translation is controlled by convergent N-methyl-D-aspartate and Gs-coupled receptor signaling pathways. *The Journal of biological chemistry*, 283(1):582–592. (Cited on page 38.)

Boccara, C. N., Sargolini, F., Thoresen, V. H., Solstad, T., Witter, M. P., Moser, E. I., and Moser, M.-B. (2010). Grid cells in pre- and parasubiculum. *Nature Neuroscience*, 13(8):987–994. (Cited on pages 42, 109, 138, 151, 210, and 256.)

Bonnevie, T., Dunn, B., Fyhn, M., Hafting, T., Derdikman, D., Kubie, J. L., Roudi, Y., Moser, E. I., and Moser, M.-B. (2013). Grid cells require excitatory drive from the hippocampus. *Nature Neuroscience*, 16(3):309–317. (Cited on pages 1, 17, and 290.)

Bozon, B., Davis, S., and Laroche, S. (2002). Regulated transcription of the immediate-early gene Zif268: mechanisms and gene dosage-dependent function in synaptic plasticity and memory formation. *Hippocampus*, 12(5):570–577. (Cited on page 33.)

Braak, H. and Braak, E. (1991). Neuropathological stageing of Alzheimer-related changes. *Acta neuropathologica*, 82(4):239–259. (Cited on pages 11, 13, 29, 30, 42, 106, 108, 153, 291, and 293.)

Braak, H., Del Tredici, K., Bohl, J., Bratzke, H., and Braak, E. (2000). Pathological changes in the parahippocampal region in select non-Alzheimer's dementias. *Annals of the New York Academy of Sciences*, 911:221–239. (Cited on pages 30, 42, and 293.)

Bramham, C. R., Alme, M. N., Bittins, M., Kuipers, S. D., Nair, R. R., Pai, B., Panja, D., Schubert, M., Soule, J., Tiron, A., and Wibrand, K. (2010). The Arc of synaptic memory. *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale*, 200(2):125–140. (Cited on pages 32, 33, 34, 37, 39, 132, 158, 160, 203, 209, 211, 215, 216, 260, 286, and 294.)

Bramham, C. R., Worley, P. F., Moore, M. J., and Guzowski, J. F. (2008). The immediate early gene arc/arg3.1: regulation, mechanisms, and function. *The Journal of Neuroscience*, 28(46):11760–11767. (Cited on page 33.)

Brun, V. H., Leutgeb, S., Wu, H.-Q., Schwarcz, R., Witter, M. P., Moser, E. I., and Moser, M.-B. (2008a). Impaired spatial representation in CA1 after lesion of direct input from entorhinal cortex. *Neuron*, 57(2):290–302. (Cited on pages 14, 16, 23, 108, and 152.)

Brun, V. H., Otnass, M. K., Molden, S., Steffenach, H.-A., Witter, M. P., Moser, M.-B., and Moser, E. I. (2002). Place cells and place recognition maintained by direct entorhinal-hippocampal circuitry. *Science (New York, NY)*, 296(5576):2243–2246. (Cited on page 14.)

Brun, V. H., Solstad, T., Kjelstrup, K. B., Fyhn, M., Witter, M. P., Moser, E. I., and Moser, M.-B. (2008b). Progressive increase in grid scale from dorsal to ventral medial entorhinal cortex. *Hippocampus*, 18(12):1200–1212. (Cited on pages 26 and 292.)

Burbach, J. P. H. and van der Zwaag, B. (2009). Contact in the genetics of autism and schizophrenia. *Trends in Neurosciences*, 32(2):69–72. (Cited on page 12.)

Burgalossi, A., Herfst, L., von Heimendahl, M., Förste, H., Haskic, K., Schmidt, M., and Brecht, M. (2011). Microcircuits of functionally identified neurons in the rat medial entorhinal cortex. *Neuron*, 70(4):773–786. (Cited on pages 20, 21, 23, 24, 27, 42, 109, 138, 140, 151, 154, 209, 210, 216, 254, 255, 256, 258, 262, and 294.)

Burgess, N., Maguire, E. A., and O'Keefe, J. (2002). The human hippocampus and spatial and episodic memory. *Neuron*, 35(4):625–641. (Cited on page 289.)

Burke, S. N., Chawla, M. K., Penner, M. R., Crowell, B. E., Worley, P. F., Barnes, C. A., and McNaughton, B. L. (2005). Differential encoding of behavior and spatial context in deep and superficial layers of the neocortex. *Neuron*, 45(5):667–674. (Cited on page 216.)

Burwell, R. D. (2000). The parahippocampal region: corticocortical connectivity. *Annals of the New York Academy of Sciences*, 911:25–42. (Cited on page 21.)

Burwell, R. D. and Amaral, D. G. (1998). Cortical afferents of the perirhinal, postrhinal, and entorhinal cortices of the rat. *The Journal of Comparative Neurology*, 398(2):179–205. (Cited on pages 209 and 286.)

Burwell, R. D. and Witter, M. P. (2002). Basic anatomy of the parahippocampal region in monkeys and rats. In Witter, M. P. and Wouterlood, F. G., editors, *The parahippocampal region*, pages 35–59. Oxford University Press. (Cited on page 19.)

Caballero-Bleda, M. and Witter, M. P. (1993). Regional and laminar organization of projections from the presubiculum and parasubiculum to the entorhinal cortex: an anterograde tracing study in the rat. *The Journal of Comparative Neurology*, 328(1):115–129. (Cited on pages 19, 20, and 151.)

Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., Xing, Y., Lubischer, J. L., Krieg, P. A., Krupenko, S. A., Thompson, W. J., and Barres, B. A. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *The Journal of Neuroscience*, 28(1):264–278. (Cited on pages 6, 7, 8, 9, 147, 148, and 153.)

Caiazzo, M., Dell'Anno, M. T., Dvoretskova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T. D., Menegon, A., Roncaglia, P., Colciago, G., Russo, G., Carninci, P., Pezzoli, G., Gainetdinov, R. R., Gustincich, S., Dityatev, A., and Broccoli, V. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature*, 476(7359):224–227. (Cited on page 9.)

Cajal, S. R. y. (1995). *Histology of the nervous system*. Oxford University Press. (Cited on page 5.)

Canto, C. B., Koganezawa, N., Beed, P., Moser, E. I., and Witter, M. P. (2012). All layers of medial entorhinal cortex receive presubicular and parasubicular inputs. *The Journal of Neuroscience*, 32(49):17620–17631. (Cited on pages 19, 25, 108, 152, 258, 267, and 286.)

Canto, C. B., Wouterlood, F. G., and Witter, M. P. (2008). What does the anatomical organization of the entorhinal cortex tell us? *Neural plasticity*, 2008:381243. (Cited on pages 13, 15, 17, 18, 20, 21, 22, 23, 24, 25, 29, 42, 100, 109, 152, 209, 257, 258, and 294.)

Catterall, W. A., Dib-Hajj, S., Meisler, M. H., and Pietrobon, D. (2008). Inherited neuronal ion channelopathies: new windows on complex neurological diseases. *The Journal of Neuroscience*, 28(46):11768–11777. (Cited on page 12.)

Cenquizca, L. A. and Swanson, L. W. (2007). Spatial organization of direct hippocampal field CA1 axonal projections to the rest of the cerebral cortex. *Brain Research reviews*, 56(1):1–26. (Cited on pages 28, 266, and 289.)

Chen, L. L., Lin, L. H., Green, E. J., Barnes, C. A., and McNaughton, B. L. (1994). Head-direction cells in the rat posterior cortex. I. Anatomical distribution and behavioral modulation. *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale*, 101(1):8–23. (Cited on page 286.)

Cherrier, M. M., Mendez, M., and Perryman, K. (2001). Route learning performance in Alzheimer disease patients. *Neuropsychiatry, neuropsychology, and behavioral neurology*, 14(3):159–168. (Cited on page 154.)

Chevaleyre, V. and Siegelbaum, S. A. (2010). Strong CA2 pyramidal neuron synapses define a powerful disynaptic cortico-hippocampal loop. *Neuron*, 66(4):560–572. (Cited on page 15.)

Chowdhury, S., Shepherd, J. D., Okuno, H., Lyford, G., Petralia, R. S., Plath, N., Kuhl, D., Huganir, R. L., and Worley, P. F. (2006). Arc/Arg3.1 interacts with the endocytic machinery to regulate AMPA receptor trafficking. *Neuron*, 52(3):445–459. (Cited on pages 38, 209, 215,

and 260.)

Christensen, D. Z., Thomsen, M. S., and Mikkelsen, J. D. (2013). Reduced basal and novelty-induced levels of activity-regulated cytoskeleton associated protein (Arc) and c-Fos mRNA in the cerebral cortex and hippocampus of APPswe/PS1ΔE9 transgenic mice. *Neurochemistry international*, 63(1):54–60. (Cited on page 256.)

Chrobak, J. J. and Buzsaki, G. (1998). Gamma oscillations in the entorhinal cortex of the freely behaving rat. *The Journal of Neuroscience*, 18(1):388–398. (Cited on page 27.)

Churchwell, J. C. and Kesner, R. P. (2011). Hippocampal-prefrontal dynamics in spatial working memory: interactions and independent parallel processing. *Behavioural brain research*, 225(2):389–395. (Cited on page 287.)

Coba, M. P., Valor, L. M., Kopanitsa, M. V., Afinowi, N. O., and Grant, S. G. N. (2008). Kinase networks integrate profiles of N-methyl-D-aspartate receptor-mediated gene expression in hippocampus. *The Journal of biological chemistry*, 283(49):34101–34107. (Cited on page 37.)

Consortium, T. E. P. (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74. (Cited on page 9.)

Conti, E. and Izaurralde, E. (2005). Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Current opinion in cell biology*, 17(3):316–325. (Cited on page 37.)

Couey, J. J., Witoelar, A., Zhang, S.-J., Zheng, K., Ye, J., Dunn, B., Czajkowski, R., Moser, M.-B., Moser, E. I., Roudi, Y., and Witter, M. P. (2013). Recurrent inhibitory circuitry as a mechanism for grid formation. *Nature Neuroscience*, pages 1–9. (Cited on pages 21, 22, 26, and 108.)

Craig, S. and Commins, S. (2006). The subiculum to entorhinal cortex projection is capable of sustaining both short- and long-term plastic changes. *Behavioural brain research*, 174(2):281–288. (Cited on pages 24, 258, 266, and 287.)

Craig, S. and Commins, S. (2007). Plastic and metaplastic changes in the CA1 and subicular projections to the entorhinal cortex. *Brain Research*, 1147:124–139. (Cited on pages 24, 258, 266, and 287.)

Croning, M. D. R., Marshall, M. C., McLaren, P., Armstrong, J. D., and Grant, S. G. N. (2009). G2Cdb: the Genes to Cognition database. *Nucleic acids research*, 37(Database):D846–D851. (Cited on page 114.)

Cunningham, M. O., Hunt, J., Middleton, S., LeBeau, F. E. N., Gillies, M. J., Gillies, M. G., Davies, C. H., Maycox, P. R., Whittington, M. A., and Racca, C. (2006). Region-specific reduction in entorhinal gamma oscillations and parvalbumin-immunoreactive neurons in animal models of psychiatric illness. *The Journal of Neuroscience*, 26(10):2767–2776. (Cited on page 31.)

Czerniawski, J., Ree, F., Chia, C., Ramamoorthi, K., Kumata, Y., and Otto, T. A. (2011). The importance of having Arc: expression of the immediate-early gene Arc is required for hippocampus-dependent fear conditioning and blocked by NMDA receptor antagonism. *The Journal of Neuroscience*, 31(31):11200–11207. (Cited on page 39.)

Davis, F. P. and Eddy, S. R. (2009). A tool for identification of genes expressed in patterns of interest using the Allen Brain Atlas. *Bioinformatics (Oxford, England)*, 25(13):1647–1654. (Cited on pages 5, 48, 53, and 55.)

de Foubert, G., O'Neill, M. J., and Zetterström, T. S. C. (2007). Acute onset by 5-HT(6)-receptor activation on rat brain brain-derived neurotrophic factor and activity-regulated cytoskeletal-associated protein mRNA expression. *Neuroscience*, 147(3):778–785. (Cited on page 32.)

de la Serna, E., Baeza, I., Toro, J., Andrés, S., Puig, O., Sánchez-Guistau, V., Romero, S., Bernardo, M., and Castro-Fornieles, J. (2009). Relationship between clinical and neuropsychological characteristics in child and adolescent first degree relatives of subjects with schizophrenia. *Schizophrenia Research*. (Cited on page 31.)

de Quervain, D. J.-F. and Papassotiropoulos, A. (2006). Identification of a genetic cluster influencing memory performance and hippocampal activity in humans. *Proceedings of the National Academy of Sciences of the United States of America*, 103(11):4270–4274. (Cited on page 32.)

Deguchi, Y., Donato, F., Galimberti, I., Cabuy, E., and Caroni, P. (2011). Temporally matched subpopulations of selectively interconnected principal neurons in the hippocampus. *Nature Neuroscience*. (Cited on page 6.)

Deller, T., Martinez, A., Nitsch, R., and Frotscher, M. (1996). A novel entorhinal projection to the rat dentate gyrus: direct innervation of proximal dendrites and cell bodies of granule cells and GABAergic neurons. *The Journal of Neuroscience*, 16(10):3322–3333. (Cited on page 25.)

Derdikman, D., Whitlock, J. R., Tsao, A., Fyhn, M., Hafting, T., Moser, M.-B., and Moser, E. I. (2009). Fragmentation of grid cell maps in a multicompartment environment. *Nature Neuroscience*, 12(10):1325–1332. (Cited on page 17.)

Devanand, D. P., Pradhaban, G., Liu, X., Khandji, A., De Santi, S., Segal, S., Rusinek, H., Pelton, G. H., Honig, L. S., Mayeux, R., Stern, Y., Tabert, M. H., and de Leon, M. J. (2007). Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology*, 68(11):828–836. (Cited on page 108.)

Dhillon, A. and Jones, R. S. (2000). Laminar differences in recurrent excitatory transmission in the rat entorhinal cortex in vitro. *Neuroscience*, 99(3):413–422. (Cited on pages 22 and 26.)

Doeller, C. F., Barry, C., and Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, 463(7281):657–661. (Cited on page 17.)

Dolorfo, C. L. and Amaral, D. G. (1998a). Entorhinal cortex of the rat: organization of intrinsic connections. *The Journal of Comparative Neurology*, 398(1):49–82. (Cited on page 21.)

Dolorfo, C. L. and Amaral, D. G. (1998b). Entorhinal cortex of the rat: topographic organization of the cells of origin of the perforant path projection to the dentate gyrus. *The Journal of Comparative Neurology*, Online(2004 Abstract Viewer and Itinerary Planner.):25–48. (Cited on pages 6, 15, 26, 28, 42, 100, 109, 209, and 266.)

Domnisoru, C., Kinkhabwala, A. A., and Tank, D. W. (2013). Membrane potential dynamics of grid cells. *Nature*. (Cited on pages 1, 23, 108, and 262.)

Dong, H.-W., Swanson, L. W., Chen, L., Fanselow, M. S., and Toga, A. W. (2009). Genomic-anatomic evidence for distinct functional domains in hippocampal field CA1. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28):11794–11799. (Cited on pages 9, 100, and 101.)

Doyle, J. P., Dougherty, J. D., Heiman, M., Schmidt, E. F., Stevens, T. R., Ma, G., Bupp, S., Shrestha, P., Shah, R. D., Doughty, M. L., Gong, S., Greengard, P., and Heintz, N. (2008). Application of a translational profiling approach for the comparative analysis of CNS cell types. *Cell*, 135(4):749–762. (Cited on pages 8, 42, and 291.)

Du, F., Whetsell, W. O., Abou-Khalil, B., Blumenkopf, B., Lothman, E. W., and Schwarcz, R. (1993). Preferential neuronal loss in layer III of the entorhinal cortex in patients with temporal lobe epilepsy. *Epilepsy research*, 16(3):223–233. (Cited on pages 13, 31, 108, 154, and 293.)

Dumitriu, D., Rodriguez, A., and Morrison, J. H. (2011). High-throughput, detailed, cell-specific neuroanatomy of dendritic spines using microinjection and confocal microscopy. *Nature protocols*, 6(9):1391–1411. (Cited on page 166.)

Duyckaerts, C., Delatour, B., and Potier, M.-C. (2009). Classification and basic pathology of Alzheimer disease. *Acta neuropathologica*, 118(1):5–36. (Cited on page 30.)

Egorov, A. V., Hamam, B. N., Fransén, E., Hasselmo, M. E., and Alonso, A. A. (2002). Graded persistent activity in entorhinal cortex neurons. *Nature*, 420(6912):173–178. (Cited on pages 16, 24, 109, 258, and 287.)

Eichenbaum, H. and Lipton, P. A. (2008). Towards a functional organization of the medial temporal lobe memory system: role of the parahippocampal and medial entorhinal cortical areas. *Hippocampus*, 18(12):1314–1324. (Cited on page 291.)

Engel, Jr, J. (1996). Introduction to temporal lobe epilepsy. *Epilepsy research*, 26(1):141–150. (Cited on page 31.)

Fanselow, M. S. and Dong, H.-W. (2010). Are the Dorsal and Ventral Hippocampus Functionally Distinct Structures? *Neuron*, 65(1):7–19. (Cited on page 28.)

Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., Myers, R. H., Pericak-Vance, M. A., Risch, N., and van Duijn, C. M. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *JAMA : the journal of the American Medical Association*, 278(16):1349–1356. (Cited on page 11.)

Flavell, S. W. and Greenberg, M. E. (2008). Signaling Mechanisms Linking Neuronal Activity to Gene Expression and Plasticity of the Nervous System. *Annual review of neuroscience*, 31:563–590. (Cited on pages 32, 132, and 156.)

Fransén, E., Alonso, A. A., Dickson, C. T., Magistretti, J., and Hasselmo, M. E. (2004). Ionic mechanisms in the generation of subthreshold oscillations and action potential clustering in entorhinal layer II stellate neurons. *Hippocampus*, 14(3):368–384. (Cited on page 109.)

Fransén, E., Tahvildari, B., Egorov, A. V., Hasselmo, M. E., and Alonso, A. A. (2006). Mechanism of graded persistent cellular activity of entorhinal cortex layer v neurons. *Neuron*, 49(5):735–746. (Cited on page 287.)

Fujimoto, T., Tanaka, H., Kumamaru, E., Okamura, K., and Miki, N. (2004). Arc interacts with microtubules/microtubule-associated protein 2 and attenuates microtubule-associated protein 2 immunoreactivity in the dendrites. *Journal of neuroscience research*, 76(1):51–63. (Cited on page 38.)

Fyhn, M., Hafting, T., Treves, A., Moser, M.-B., and Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, 446(7132):190–194. (Cited on pages 17, 215, 287, and 293.)

Fyhn, M., Hafting, T., Witter, M. P., Moser, E. I., and Moser, M.-B. (2008). Grid cells in mice. *Hippocampus*, 18(12):1230–1238. (Cited on pages 17, 26, 100, 208, and 255.)

Fyhn, M., Molden, S., Witter, M. P., Moser, E. I., and Moser, M.-B. (2004). Spatial representation in the entorhinal cortex. *Science (New York, NY)*, 305(5688):1258–1264. (Cited on pages 1, 17, and 26.)

Gao, M., Sossa, K., Song, L., Errington, L., Cummings, L., Hwang, H., Kuhl, D., Worley, P., and Lee, H.-K. (2010). A specific requirement of Arc/Arg3.1 for visual experience-induced homeostatic synaptic plasticity in mouse primary visual cortex. *The Journal of Neuroscience*, 30(21):7168–7178. (Cited on page 38.)

Gao, Y., Tatavarty, V., Korza, G., Levin, M. K., and Carson, J. H. (2008). Multiplexed dendritic targeting of alpha calcium calmodulin-dependent protein kinase II, neurogranin, and activity-regulated cytoskeleton-associated protein RNAs by the A2 pathway. *Molecular biology of the cell*, 19(5):2311–2327. (Cited on page 34.)

Garden, D. L. F., Dodson, P. D., O'Donnell, C., White, M. D., and Nolan, M. F. (2008). Tuning of synaptic integration in the medial entorhinal cortex to the organization of grid cell firing fields. *Neuron*, 60(5):875–889. (Cited on pages 9, 13, 20, 21, 26, 27, 42, 100, 103, 109, 129, 138, 155, 158, 208, 209, 256, 292, and 293.)

Gasparini, S. (2011). Distance- and activity-dependent modulation of spike back-propagation in layer V pyramidal neurons of the medial entorhinal cortex. *Journal of Neurophysiology*, 105(3):1372–1379. (Cited on pages 25 and 286.)

Gatome, C. W., Slomianka, L., Lipp, H. P., and Amrein, I. (2010). Number estimates of neuronal phenotypes in layer II of the medial entorhinal cortex of rat and mouse. *Neuroscience*, 170(1):156–165. (Cited on pages 21, 27, 100, 103, 177, and 255.)

Geschwind, D. H. and Konopka, G. (2009). Neuroscience in the era of functional genomics and systems biology. *Nature*, 461(7266):908–915. (Cited on page 7.)

Giocomo, L. M. and Hasselmo, M. E. (2008). Time constants of h current in layer ii stellate cells differ along the dorsal to ventral axis of medial entorhinal cortex. *The Journal of Neuroscience*, 28(38):9414–9425. (Cited on pages 13 and 20.)

Giocomo, L. M. and Hasselmo, M. E. (2009). Knock-out of HCN1 subunit flattens dorsal-ventral frequency gradient of medial entorhinal neurons in adult mice. *The Journal of Neuroscience*, 29(23):7625–7630. (Cited on page 138.)

Giorgi, C., Yeo, G. W., Stone, M. E., Katz, D. B., Burge, C., Turrigiano, G., and Moore, M. J. (2007). The EJC factor eIF4AIII modulates synaptic strength and neuronal protein expression. *Cell*, 130(1):179–191. (Cited on pages 37 and 158.)

Goff, D. C. and Coyle, J. T. (2001). The emerging role of glutamate in the pathophysiology and treatment of schizophrenia. *The American journal of psychiatry*, 158(9):1367–1377. (Cited on page 12.)

Gómez-Isla, T., Price, J. L., McKeel, D. W., Morris, J. C., Growdon, J. H., and Hyman, B. T. (1996). Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *The Journal of Neuroscience*, 16(14):4491–4500. (Cited on pages 30 and 291.)

Gothard, K. M., Skaggs, W. E., and McNaughton, B. L. (1996a). Dynamics of mismatch correction in the hippocampal ensemble code for space: interaction between path integration and environmental cues. *The Journal of Neuroscience*, 16(24):8027–8040. (Cited on page 14.)

Gothard, K. M., Skaggs, W. E., Moore, K. M., and McNaughton, B. L. (1996b). Binding of hippocampal CA1 neural activity to multiple reference frames in a landmark-based navigation task. *The Journal of Neuroscience*, 16(2):823–835. (Cited on page 256.)

Grant, S. G. N. (2005). Synapse proteomics of multiprotein complexes: en route from genes to nervous system diseases. *Human Molecular Genetics*, 14(suppl 2):R225–R234. (Cited on page 10.)

Gusev, P. A., Cui, C., Alkon, D. L., and Gubin, A. N. (2005). Topography of Arc/Arg3.1 mRNA expression in the dorsal and ventral hippocampus induced by recent and remote spatial memory recall: dissociation of CA3 and CA1 activation. *The Journal of Neuroscience*, 25(41):9384–9397. (Cited on pages 35, 36, 210, 215, 216, 217, 256, and 294.)

Guzowski, J. F., McNaughton, B. L., Barnes, C. A., and Worley, P. F. (1999). Environment-specific expression of the immediate-early gene Arc in hippocampal neuronal ensembles. *Nature Neuroscience*, 2(12):1120–1124. (Cited on pages 34, 35, 215, 236, 254, 256, 257, and 261.)

Guzowski, J. F., Miyashita, T., Chawla, M. K., Sanderson, J., Maes, L. I., Houston, F. P., Lipa, P., McNaughton, B. L., Worley, P. F., and Barnes, C. A. (2006). Recent behavioral history modifies coupling between cell activity and Arc gene transcription in hippocampal CA1 neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4):1077–1082. (Cited on pages 34, 35, 36, 215, 255, 256, and 259.)

Guzowski, J. F., Setlow, B., Wagner, E. K., and McGaugh, J. L. (2001). Experience-dependent gene expression in the rat hippocampus after spatial learning: a comparison of the immediate-early genes Arc, c-fos, and zif268. *The Journal of Neuroscience*, 21(14):5089–5098. (Cited on pages 34, 35, 36, 38, 216, 217, and 261.)

Hafting, T., Fyhn, M., Bonnevie, T., Moser, M.-B., and Moser, E. I. (2008). Hippocampus-independent phase precession in entorhinal grid cells. *Nature*, 453(7199):1248–1252. (Cited on pages 109 and 208.)

Hafting, T., Fyhn, M., Molden, S., Moser, M.-B., and Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, 436(7052):801–806. (Cited on pages 1, 9, 13, 17, 26, 42, 100, 129, 158, 159, 215, 216, 254, 255, 256, 258, 262, 266, 291, 292, and 293.)

Hamam, B. N., Amaral, D. G., and Alonso, A. A. (2002). Morphological and electrophysiological characteristics of layer V neurons of the rat lateral entorhinal cortex. *The Journal of Comparative Neurology*, 451(1):45–61. (Cited on page 286.)

Hamam, B. N., Kennedy, T. E., Alonso, A., and Amaral, D. G. (2000). Morphological and electrophysiological characteristics of layer V neurons of the rat medial entorhinal cortex. *The Journal of Comparative Neurology*, 418(4):457–472. (Cited on pages 25, 109, 258, and 266.)

Hargreaves, E. L., Rao, G., Lee, I., and Knierim, J. J. (2005). Major dissociation between medial and lateral entorhinal input to dorsal hippocampus. *Science (New York, NY)*, 308(5729):1792–1794. (Cited on page 13.)

Harich, S., Kinfe, T., Koch, M., and Schwabe, K. (2008). Neonatal lesions of the entorhinal cortex induce long-term changes of limbic brain regions and maze learning deficits in adult rats. *Neuroscience*, 153(4):918–928. (Cited on page 16.)

Harmar, A. J., Hills, R. A., Rosser, E. M., Jones, M., Buneman, O. P., Dunbar, D. R., Greenhill, S. D., Hale, V. A., Sharman, J. L., Bonner, T. I., Catterall, W. A., Davenport, A. P., Delagrange, P., Dollery, C. T., Foord, S. M., Gutman, G. A., Laudet, V., Neubig, R. R., Ohlstein, E. H., Olsen, R. W., Peters, J., Pin, J.-P., Ruffolo, R. R., Searls, D. B., Wright, M. W., and Spedding, M. (2009). IUPHAR-DB: the IUPHAR database of G protein-coupled receptors and ion channels. *Nucleic acids research*, 37(Database issue):D680–5. (Cited on page 114.)

Harris, J. A., Devidze, N., Verret, L., Ho, K., Halabisky, B., Thwin, M. T., Kim, D., Hamto, P., Lo, I., Yu, G.-Q., Palop, J. J., Masliah, E., and Mucke, L. (2010). Transsynaptic progression of amyloid-β-induced neuronal dysfunction within the entorhinal-hippocampal network. *Neuron*, 68(3):428–441. (Cited on page 30.)

Harrison, P. J. and Weinberger, D. R. (2005). Schizophrenia genes, gene expression, and neuropathology: on the matter of their convergence. *Molecular psychiatry*, 10(1):40–68– image 5. (Cited on page 12.)

Hartzell, A. L., Burke, S. N., Hoang, L. T., Lister, J. P., Rodriguez, C. N., and Barnes, C. A. (2013). Transcription of the Immediate-Early Gene Arc in CA1 of the Hippocampus Reveals Activity Differences along the Proximodistal Axis That Are Attenuated by Advanced Age. *The Journal of Neuroscience*, 33(8):3424–3433. (Cited on pages 15, 35, and 36.)

Hawrylycz, M., Baldock, R. A., Burger, A., Hashikawa, T., Johnson, G. A., Martone, M., Ng, L., Lau, C., Larson, S. D., Larsen, S. D., Nissanov, J., Puelles, L., Ruffins, S., Verbeek, F., Zaslavsky, I., and Boline, J. (2011a). Digital atlasing and standardization in the mouse brain. *PLoS Computational Biology*, 7(2):e1001065. (Cited on page 5.)

Hawrylycz, M., Bernard, A., Lau, C., Sunkin, S. M., Chakravarty, M. M., Lein, E. S., Jones, A. R., and Ng, L. (2010). Areal and laminar differentiation in the mouse neocortex using large scale gene expression data. *Methods (San Diego, Calif)*, 50(2):113–121. (Cited on page 5.)

Hawrylycz, M., Ng, L., Page, D., Morris, J., Lau, C., Faber, S., Faber, V., Sunkin, S., Menon, V., Lein, E., and Jones, A. (2011b). Multi-scale correlation structure of gene expression in the brain. *Neural networks*, 24(9):933–942. (Cited on page 5.)

Heiman, M., Schaefer, A., Gong, S., Peterson, J. D., Day, M., Ramsey, K. E., Suárez-Fariñas, M., Schwarz, C., Stephan, D. A., Surmeier, D. J., Greengard, P., and Heintz, N. (2008). A translational profiling approach for the molecular characterization of CNS cell types. *Cell*, 135(4):738–748. (Cited on page 8.)

Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). *Nature Neuroscience*, 7(5):483. (Cited on page 5.)

Hemby, S. E., Ginsberg, S. D., Brunk, B., Arnold, S. E., Trojanowski, J. Q., and Eberwine, J. H. (2002). Gene expression profile for schizophrenia: discrete neuron transcription patterns in the entorhinal cortex. *Archives of General Psychiatry*, 59(7):631–640. (Cited on page 30.)

Henriksen, E. J., Colgin, L. L., Barnes, C. A., Witter, M. P., Moser, M.-B., and Moser, E. I. (2010). Spatial representation along the proximodistal axis of CA1. *Neuron*, 68(1):127–137. (Cited on pages 15 and 36.)

Hobert, O., Carrera, I., and Stefanakis, N. (2010). The molecular and gene regulatory signature of a neuron. *Trends in Neurosciences*, 33(10):435–445. (Cited on page 9.)

Hock, B. J. J. and Bunsey, M. D. (1998). Differential Effects of Dorsal and Ventral Hippocampal Lesions. *The Journal of Neuroscience*, 18(17):7027–7032. (Cited on page 28.)

Holloway, C. M. and McIntyre, C. K. (2011). Post-training disruption of Arc protein expression in the anterior cingulate cortex impairs long-term memory for inhibitory avoidance training. *Neurobiology of learning and memory*, 95(4):425–432. (Cited on page 39.)

Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols*, 4(1):44–57. (Cited on pages 112, 113, and 114.)

Huang, Y. and Mucke, L. (2012). Alzheimer Mechanisms and Therapeutic Strategies. *Cell*, 148(6):1204–1222. (Cited on pages 10 and 11.)

Ikeda, J., Mori, K., Oka, S., and Watanbe, Y. (1989). A columnar arrangement of dendritic processes of entorhinal cortex neurons revealed by a monoclonal antibody. *Brain Research*, 505(1):176–179. (Cited on page 19.)

Insausti, R., Herrero, M. T., and Witter, M. P. (1997). Entorhinal cortex of the rat: cytoarchitectonic subdivisions and the origin and distribution of cortical efferents. *Hippocampus*, 7(2):146–183. (Cited on pages 18, 20, 23, 25, 108, 109, 144, 170, 216, 230, 254, 255, 266, 267, 286, and 287.)

Jarrard, L. E. (1989). On the use of ibotenic acid to lesion selectively different components of the hippocampal formation. *Journal of neuroscience methods*, 29(3):251–259. (Cited on pages 267, 273, and 274.)

Jarrard, L. E., Luu, L. P., and Davidson, T. L. (2012). A study of hippocampal structure-function relations along the septo-temporal axis. *Hippocampus*, 22(4):680–692. (Cited on pages 28 and 271.)

Jeffery, K. J. (2007). Self-localization and the entorhinal-hippocampal system. *Current opinion in neurobiology*, 17(6):684–691. (Cited on pages 13 and 17.)

Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., and Lempicki, R. A. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics (Oxford, England)*, 28:1805–1806. (Cited on page 113.)

Jones, A. R., Overly, C. C., and Sunkin, S. M. (2009). The Allen Brain Atlas: 5 years and beyond. *Nature Reviews Neuroscience*, 10(11):821–828. (Cited on page 7.)

Jones, B. F. and Witter, M. P. (2007). Cingulate cortex projections to the parahippocampal region and hippocampal formation in the rat. *Hippocampus*, 17(10):957–976. (Cited on pages 19, 24, 258, and 266.)

Jones, R. S. and Buhl, E. H. (1993). Basket-like interneurones in layer II of the entorhinal cortex exhibit a powerful NMDA-mediated synaptic excitation. *Neuroscience letters*, 149(1):35–39. (Cited on pages 22 and 255.)

Jung, M. W., Wiener, S. I., and McNaughton, B. L. (1994). Comparison of spatial firing characteristics of units in dorsal and ventral hippocampus of the rat. *The Journal of Neuroscience*, 14(12):7347–7356. (Cited on page 27.)

Kajiwara, R., Wouterlood, F. G., Sah, A., Boekel, A. J., Baks-te Bulte, L. T. G., and Witter, M. P. (2008). Convergence of entorhinal and CA3 inputs onto pyramidal neurons and interneurons in hippocampal area CA1–an anatomical study in the rat. *Hippocampus*, 18(3):266–280. (Cited on page 23.)

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic acids research*, 32(Database issue):D277–80. (Cited on pages 11, 113, 115, 123, 128, 129, 132, 136, and 154.)

Kasperaviciute, D., Catarino, C. B., Heinzen, E. L., Depondt, C., Cavalleri, G. L., Caboclo, L. O., Tate, S. K., Jamnadas-Khoda, J., Chinthapalli, K., Clayton, L. M. S., Shianna, K. V., Radtke, R. A., Mikati, M. A., Gallentine, W. B., Husain, A. M., Alhusaini, S., Leppert, D., Middleton, L. T., Gibson, R. A., Johnson, M. R., Matthews, P. M., Hosford, D., Heuser, K., Amos, L., Ortega, M., Zumsteg, D., Wieser, H. G., Steinhoff, B. J., Kramer, G., Hansen, J., Dorn, T.,

Kantanen, A. M., Gjerstad, L., Peuralinna, T., Hernandez, D. G., Eriksson, K. J., Kalviainen, R. K., Doherty, C. P., Wood, N. W., Pandolfo, M., Duncan, J. S., Sander, J. W., Delanty, N., Goldstein, D. B., and Sisodiya, S. M. (2010). Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study. *Brain*, 133(7):2136–2147. (Cited on page 12.)

Kawashima, T., Okuno, H., Nonaka, M., Adachi-Morishima, A., Kyo, N., Okamura, M., Takemoto-Kimura, S., Worley, P. F., and Bito, H. (2009). Synaptic activity-responsive element in the Arc/Arg3.1 promoter essential for synapse-to-nucleus signaling in activated neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 106(1):316–321. (Cited on pages 33, 37, and 132.)

Kelly, M. P. and Deadwyler, S. A. (2002). Acquisition of a novel behavior induces higher levels of Arc mRNA than does overtrained performance. *Neuroscience*, 110(4):617–626. (Cited on pages 35 and 36.)

Kelly, M. P. and Deadwyler, S. A. (2003). Experience-dependent regulation of the immediate-early gene arc differs across brain regions. *The Journal of Neuroscience*, 23(16):6443–6451. (Cited on pages 34, 36, and 255.)

Kentros, C., Hargreaves, E., Hawkins, R. D., Kandel, E. R., Shapiro, M., and Muller, R. V. (1998). Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science (New York, NY)*, 280(5372):2121–2126. (Cited on page 14.)

Kentros, C. G., Agnihotri, N. T., Streater, S., Hawkins, R. D., and Kandel, E. R. (2004). Increased attention to spatial context increases both place field stability and spatial memory. *Neuron*, 42(2):283–295. (Cited on page 257.)

Kim, Y.-A., Wuchty, S., and Przytycka, T. M. (2011). Identifying causal genes and dysregulated pathways in complex diseases. *PLoS Computational Biology*, 7(3):e1001095. (Cited on page 12.)

Kirsch, L., Liscovitch, N., and Chechik, G. (2012). Localizing Genes to Cerebellar Layers by Classifying ISH Images. *PLoS Computational Biology*, 8(12):e1002790. (Cited on pages 4, 8, 43, and 101.)

Kjelstrup, K. G., Tuvnes, F. A., Steffenach, H.-A., Murison, R., Moser, E. I., and Moser, M.-B. (2002). Reduced fear expression after lesions of the ventral hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 99(16):10825–10830. (Cited on pages 28 and 289.)

Kjonigsen, L. J., Leergaard, T. B., Witter, M. P., and Bjaalie, J. G. (2011). Digital Atlas of Anatomical Subdivisions and Boundaries of the Rat Hippocampal Region. *Frontiers in Neuroinformatics*, 5. (Cited on pages 15 and 18.)

Klassen, T., Davis, C., Goldman, A., Burgess, D., Chen, T., Wheeler, D., McPherson, J., Bourquin, T., Lewis, L., Villasana, D., Morgan, M., Muzny, D., Gibbs, R., and Noebels, J. (2011). Exome Sequencing of Ion Channel Genes Reveals Complex Profiles Confounding Personal Risk Assessment in Epilepsy. *Cell*, 145(7):1036–1048. (Cited on page 12.)

Klink, R. and Alonso, A. (1997). Morphological characteristics of layer II projection neurons in the rat medial entorhinal cortex. *Hippocampus*, 7(5):571–583. (Cited on pages 21 and 23.)

Kloosterman, F., Witter, M. P., and van Haeften, T. (2003). Topographical and laminar organization of subicular projections to the parahippocampal region of the rat. *The Journal of Comparative Neurology*, 455(2):156–171. (Cited on pages 15, 24, 258, and 266.)

Klur, S., Muller, C., Pereira de Vasconcelos, A., Ballard, T., Lopez, J., Galani, R., Certa, U., and Cassel, J.-C. (2009). Hippocampal-dependent spatial memory functions might be lateralized in rats: An approach combining gene expression profiling and reversible inactivation. *Hippocampus*, 19(9):800–816. (Cited on pages 10, 32, and 289.)

Ko, Y., Ament, S. A., Eddy, J. A., Caballero, J., Earls, J. C., Hood, L., and Price, N. D. (2013). Cell type-specific genes show striking and distinct patterns of spatial expression in the mouse brain. *Proceedings of the National Academy of Sciences of the United States of America*, 110(8):3095–3100. (Cited on pages 46 and 153.)

Kobayashi, H., Yamamoto, S., Maruo, T., and Murakami, F. (2005). Identification of a cis-acting element required for dendritic targeting of activity-regulated cytoskeleton-associated protein mRNA. *The European Journal of Neuroscience*, 22(12):2977–2984. (Cited on pages 33, 158, 161, and 260.)

Koganezawa, N., Taguchi, A., Tominaga, T., Ohara, S., Tsutsui, K.-I., Witter, M. P., and Iijima, T. (2008). Significance of the deep layers of entorhinal cortex for transfer of both perirhinal and amygdala inputs to the hippocampus. *Neuroscience research*, 61(2):172–181. (Cited on page 25.)

Köhler, C., Eriksson, L., Davies, S., and Chan-Palay, V. (1986). Neuropeptide Y innervation of the hippocampal region in the rat and monkey brain. *The Journal of Comparative Neurology*, 244(3):384–400. (Cited on page 19.)

Korb, E. and Finkbeiner, S. (2011). Arc in synaptic plasticity: from gene to behavior. *Trends in Neurosciences*, 34(11):591–598. (Cited on page 38.)

Kubik, S., Miyashita, T., Kubik-Zahorodna, A., and Guzowski, J. F. (2012). Loss of activity-dependent Arc gene expression in the retrosplenial cortex after hippocampal inactivation: interaction in a higher-order memory circuit. *Neurobiology of learning and memory*, 97(1):124–131. (Cited on pages 210 and 288.)

Kuipers, S. D., Tiron, A., Soule, J., Messaoudi, E., Trentani, A., and Bramham, C. R. (2009). Selective survival and maturation of adult-born dentate granule cells expressing the immediate early gene Arc/Arg3.1. *PloS one*, 4(3):e4885. (Cited on page 35.)

Kumar, S. S. and Buckmaster, P. S. (2006). Hyperexcitability, interneurons, and loss of GABAergic synapses in entorhinal cortex in a model of temporal lobe epilepsy. *The Journal of Neuroscience*, 26(17):4613–4623. (Cited on pages 152 and 154.)

Kumar, S. S., Jin, X., Buckmaster, P. S., and Huguenard, J. R. (2007). Recurrent circuits in layer II of medial entorhinal cortex in a model of temporal lobe epilepsy. *The Journal of Neuroscience*, 27(6):1239–1246. (Cited on pages 22 and 31.)

Langston, R. F., Ainge, J. A., Couey, J. J., Canto, C. B., Bjerknes, T. L., Witter, M. P., Moser, E. I., and Moser, M.-B. (2010). Development of the spatial representation system in the rat. *Science (New York, NY)*, 328(5985):1576–1580. (Cited on page 258.)

Langston, R. F. and Wood, E. R. (2010). Associative recognition and the hippocampus: differential effects of hippocampal lesions on object-place, object-context and object-place-context memory. *Hippocampus*, 20(10):1139–1153. (Cited on pages 287 and 288.)

Lau, C., Ng, L., Thompson, C., Pathak, S., Kuan, L., Jones, A., and Hawrylycz, M. (2008). Exploration and visualization of gene expression with neuroanatomy in the adult mouse brain. *BMC Bioinformatics*, 9:153. (Cited on page 54.)

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., Boe, A. F., Boguski, M. S., Brockway, K. S., Byrnes, E. J., Chen, L., Chen, L., Chen, T.-M., Chin, M. C., Chong, J., Crook, B. E., Czaplinska, A., Dang, C. N., Datta, S., Dee, N. R., Desaki, A. L., Desta, T., Diep, E., Dolbeare, T. A., Donelan, M. J., Dong, H.-W., Dougherty, J. G., Duncan, B. J., Ebbert, A. J., Eichele, G., Estin, L. K., Faber, C., Facer, B. A., Fields, R., Fischer, S. R., Fliss, T. P., Frensley, C., Gates, S. N., Glattfelder, K. J., Halverson, K. R., Hart, M. R., Hohmann, J. G., Howell, M. P., Jeung, D. P., Johnson, R. A., Karr, P. T., Kawal, R., Kidney, J. M., Knapik, R. H., Kuan, C. L., Lake, J. H., Laramee, A. R., Larsen, K. D., Lau, C., Lemon, T. A., Liang, A. J., Liu, Y., Luong, L. T., Michaels, J., Morgan, J. J., Morgan, R. J., Mortrud, M. T., Mosqueda, N. F., Ng, L. L., Ng, R., Orta, G. J., Overly, C. C., Pak, T. H., Parry, S. E., Pathak, S. D., Pearson, O. C., Puchalski, R. B., Riley, Z. L., Rockett, H. R., Rowland, S. A., Royall, J. J., Ruiz, M. J., Sarno, N. R., Schaffnit, K., Shapovalova, N. V., Sivisay, T., Slaughterbeck, C. R., Smith, S. C., Smith, K. A., Smith, B. I., Sodt, A. J., Stewart, N. N., Stumpf, K.-R., Sunkin, S. M., Sutram, M., Tam, A., Teemer, C. D., Thaller, C., Thompson, C. L., Varnam, L. R., Visel, A., Whitlock, R. M., Wohnoutka, P. E., Wolkey, C. K., Wong, V. Y., Wood, M., Yaylaoglu, M. B., Young, R. C., Youngstrom, B. L., Yuan, X. F., Zhang, B., Zwingman, T. A., and Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, 445(7124):168–176. (Cited on pages 4, 5, 7, 43, 49, 150, 208, and 210.)

Lesage, S. and Brice, A. (2009). Parkinson's disease: from monogenic forms to genetic susceptibility factors. *Human Molecular Genetics*, 18(1). (Cited on page 11.)

Leslie, J. H. and Nedivi, E. (2011). Activity-regulated genes as mediators of neural circuit plasticity. *Progress in Neurobiology*, 94(3):223–237. (Cited on pages 32, 33, and 292.)

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., and Moser, M.-B. (2005). Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science (New York, NY)*, 309(5734):619–623. (Cited on page 14.)

Lichtman, J. W. and Denk, W. (2011). The big and the small: challenges of imaging the brain's circuits. *Science (New York, NY)*, 334(6056):618–623. (Cited on pages 5 and 6.)

Lichtman, J. W. and Sanes, J. R. (2008). Ome sweet ome: what can the genome tell us about the connectome? *Current opinion in neurobiology*, 18(3):346–353. (Cited on page 5.)

Lim, B. K., Cho, S.-j., Sumbre, G., and Poo, M.-m. (2010). Region-specific contribution of ephrin-B and Wnt signaling to receptive field plasticity in developing optic tectum. *Neuron*, 65(6):899–911. (Cited on page 9.)

Lim, J. S. (1990). Two-dimensional signal and image processing. *Englewood Cliffs, NJ, Prentice Hall, 1990, 710 p.*, 1. (Cited on page 173.)

Lin, L., Lesnick, T. G., Maraganore, D. M., and Isacson, O. (2009). Axon guidance and synaptic maintenance: preclinical markers for neurodegenerative disease and therapeutics. *Trends in Neurosciences*, 32(3):142–149. (Cited on page 11.)

Lingenhöhl, K. and Finch, D. M. (1991). Morphological characterization of rat entorhinal neurons in vivo: soma-dendritic structure and axonal domains. *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale*, 84(1):57–74. (Cited on page 21.)

Link, W., Konietzko, U., Kauselmann, G., Krug, M., Schwanke, B., Frey, U., and Kuhl, D. (1995). Somatodendritic expression of an immediate early gene is regulated by synaptic activity. *Proceedings of the National Academy of Sciences of the United States of America*, 92(12):5734–5738. (Cited on pages 33 and 37.)

Livet, J., Weissman, T. A., Kang, H., Draft, R. W., Lu, J., Bennis, R. A., Sanes, J. R., and Lichtman, J. W. (2007). Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, 450(7166):56–62. (Cited on page 6.)

Luo, L., Callaway, E. M., and Svoboda, K. (2008). Genetic dissection of neural circuits. *Neuron*, 57(5):634–660. (Cited on page 7.)

Lyford, G. L., Yamagata, K., Kaufmann, W. E., Barnes, C. A., Sanders, L. K., Copeland, N. G., Gilbert, D. J., Jenkins, N. A., Lanahan, A. A., and Worley, P. F. (1995). Arc, a growth factor and activity-regulated gene, encodes a novel cytoskeleton-associated protein that is enriched in neuronal dendrites. *Neuron*, 14(2):433–445. (Cited on pages 32, 33, 37, 38, 209, 210, 213, 215, 219, 254, 255, and 286.)

Maddox, S. A. and Schafe, G. E. (2011). The activity-regulated cytoskeletal-associated protein (Arc/Arg3.1) is required for reconsolidation of a Pavlovian fear memory. *The Journal of Neuroscience*, 31(19):7073–7082. (Cited on page 39.)

Mamiya, N., Fukushima, H., Suzuki, A., Matsuyama, Z., Homma, S., Frankland, P. W., and Kida, S. (2009). Brain region-specific gene expression activation required for reconsolidation and extinction of contextual fear memory. *The Journal of Neuroscience*, 29(2):402–413. (Cited on page 39.)

Masdeu, J. C., Zubieta, J. L., and Arbizu, J. (2005). Neuroimaging as a marker of the onset and progression of Alzheimer's disease. *Journal of the neurological sciences*, 236(1-2):55–64. (Cited on page 30.)

McCurry, C. L., Shepherd, J. D., Tropea, D., Wang, K. H., Bear, M. F., and Sur, M. (2010). Loss of Arc renders the visual cortex impervious to the effects of sensory experience or deprivation. *Nature Neuroscience*, 13(4):450–457. (Cited on page 39.)

McDonald, A. J. and Mascagni, F. (1997). Projections of the lateral entorhinal cortex to the amygdala: a Phaseolus vulgaris leucoagglutinin study in the rat. *Neuroscience*, 77(2):445–

459. (Cited on page 25.)

McIntyre, C. K., Miyashita, T., Setlow, B., Marjon, K. D., Steward, O., Guzowski, J. F., and McGaugh, J. L. (2005). Memory-influencing intra-basolateral amygdala drug infusions modulate expression of Arc protein in the hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 102(30):10718–10723. (Cited on page 39.)

McNaughton, B. L., Barnes, C. A., Meltzer, J., and Sutherland, R. J. (1989). Hippocampal granule cells are necessary for normal spatial learning but not for spatially-selective pyramidal cell discharge. *Experimental brain research Experimentelle Hirnforschung Expérimentation cérébrale*, 76(3):485–496. (Cited on page 14.)

McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I., and Moser, M.-B. (2006a). Path integration and the neural basis of the 'cognitive map'. *Nature Reviews Neuroscience*, 7(8):663–678. (Cited on pages 14 and 266.)

McNaughton, N., Ruan, M., and Woodnorth, M.-A. (2006b). Restoring theta-like rhythmicity in rats restores initial learning in the Morris water maze. *Hippocampus*, 16(12):1102–1110. (Cited on page 36.)

Medinilla, V., Johnson, O., and Gasparini, S. (2013). Features of proximal and distal excitatory synaptic inputs to layer V neurons of the rat medial entorhinal cortex. *The Journal of Physiology*, 591(Pt 1):169–183. (Cited on pages 25, 258, 266, and 286.)

Meijering, E. (2010). Neuron tracing in perspective. *Cytometry Part A*, 77A(7):693–704. (Cited on page 173.)

Meijering, E. (2013). FeatureJ. (Cited on page 60.)

Melzer, S., Michael, M., Caputi, A., Eliava, M., Fuchs, E. C., Whittington, M. A., and Monyer, H. (2012). Long-range-projecting GABAergic neurons modulate inhibition in hippocampus and entorhinal cortex. *Science (New York, NY)*, 335(6075):1506–1510. (Cited on pages 15 and 24.)

Messaoudi, E., Kanhema, T., Soule, J., Tiron, A., Dagyte, G., da Silva, B., and Bramham, C. R. (2007). Sustained Arc/Arg3.1 synthesis controls long-term potentiation consolidation through regulation of local actin polymerization in the dentate gyrus in vivo. *The Journal of Neuroscience*, 27(39):10445–10455. (Cited on pages 38, 158, 159, 210, 215, 219, and 236.)

Miyashita, T., Kubik, S., Haghighi, N., Steward, O., and Guzowski, J. F. (2009). Rapid activation of plasticity-associated gene transcription in hippocampal neurons provides a mechanism for encoding of one-trial experience. *The Journal of Neuroscience*, 29(4):898–906. (Cited on pages 10, 32, 36, and 259.)

Mizumori, S. J., McNaughton, B. L., Barnes, C. A., and Fox, K. B. (1989). Preserved spatial coding in hippocampal CA1 pyramidal cells during reversible suppression of CA3c output: evidence for pattern completion in hippocampus. *The Journal of Neuroscience*, 9(11):3915–3928. (Cited on page 36.)

modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. A., Ma, L., Lin, M. F., Washietl, S., Arshinoff, B. I., Ay, F., Meyer, P. E., Robine, N., Washington, N. L., Di Stefano, L., Berezikov, E., Brown, C. D., Candeias, R., Carlson, J. W., Carr, A., Jungreis, I., Marbach, D., Sealfon, R., Tolstorukov, M. Y., Will, S., Alekseyenko, A. A., Artieri, C., Booth, B. W., Brooks, A. N., Dai, Q., Davis, C. A., Duff, M. O., Feng, X., Gorchakov, A. A., Gu, T., Henikoff, J. G., Kapranov, P., Li, R., MacAlpine, H. K., Malone, J., Minoda, A., Nordman, J., Okamura, K., Perry, M., Powell, S. K., Riddle, N. C., Sakai, A., Samsonova, A., Sandler, J. E., Schwartz, Y. B., Sher, N., Spokony, R., Sturgill, D., van Baren, M., Wan, K. H., Yang, L., Yu, C., Feingold, E., Good, P., Guyer, M., Lowdon, R., Ahmad, K., Andrews, J., Berger, B., Brenner, S. E., Brent, M. R., Cherbas, L., Elgin, S. C. R., Gingeras, T. R., Grossman, R., Hoskins, R. A., Kaufman, T. C., Kent, W., Kuroda, M. I., Orr-Weaver, T., Perrimon, N., Pirrotta, V., Posakony, J. W., Ren, B., Russell, S., Cherbas, P., Graveley, B. R., Lewis, S., Micklem, G., Oliver, B., Park, P. J., Celniker, S. E., Henikoff, S., Karpen, G. H., Lai, E. C., MacAlpine, D. M., Stein, L. D., White, K. P., and Kellis, M. (2010). Identification of functional elements and regulatory circuits by

Drosophila modENCODE. *Science (New York, NY)*, 330(6012):1787–1797. (Cited on page 9.)

Morris, R. G., Garrud, P., Rawlins, J. N., and O'Keefe, J. (1982). Place navigation impaired in rats with hippocampal lesions. *Nature*, 297(5868):681–683. (Cited on pages 14, 266, and 287.)

Moser, E., Moser, M.-B., and Andersen, P. (1993). Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. *The Journal of Neuroscience*, 13(9):3916–3925. (Cited on pages 14, 266, 287, and 289.)

Moser, E. I., Kropff, E., and Moser, M.-B. (2008). Place cells, grid cells, and the brain's spatial representation system. *Annual review of neuroscience*, 31:69–89. (Cited on pages 13, 14, 16, 17, 108, and 215.)

Mouse Genome Sequencing Consortium, Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigó, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., Karolchik, D., Kasprzyk, A., Kawai, J., Keibler, E., Kells, C., Kent, W. J., Kirby, A., Kolbe, D. L., Korf, I., Kucherlapati, R. S., Kulbokas, E. J., Kulp, D., Landers, T., Leger, J. P., Leonard, S., Letunic, I., Levine, R., Li, J., Li, M., Lloyd, C., Lucas, S., Ma, B., Maglott, D. R., Mardis, E. R., Matthews, L., Mauceli, E., Mayer, J. H., McCarthy, M., McCombie, W. R., McLaren, S., McLay, K., McPherson, J. D., Meldrim, J., Meredith, B., Mesirov, J. P., Miller, W., Miner, T. L., Mongin, E., Montgomery, K. T., Morgan, M., Mott, R., Mullikin, J. C., Muzny, D. M., Nash, W. E., Nelson, J. O., Nhan, M. N., Nicol, R., Ning, Z., Nusbaum, C., O'Connor, M. J., Okazaki, Y., Oliver, K., Overton-Larty, E., Pachter, L., Parra, G., Pepin, K. H., Peterson, J., Pevzner, P., Plumb, R., Pohl, C. S., Poliakov, A., Ponce, T. C., Ponting, C. P., Potter, S., Quail, M., Reymond, A., Roe, B. A., Roskin, K. M., Rubin, E. M., Rust, A. G., Santos, R., Sapojnikov, V., Schultz, B., Schultz, J., Schwartz, M. S., Schwartz, S., Scott, C., Seaman, S., Searle, S., Sharpe, T., Sheridan, A., Shownkeen, R., Sims, S., Singer, J. B., Slater, G., Smit, A., Smith, D. R., Spencer, B., Stabenau, A., Stange-Thomann, N., Sugnet, C., Suyama, M., Tesler, G., Thompson, J., Torrents, D., Trevaskis, E., Tromp, J., Ucla, C., Ureta-Vidal, A., Vinson, J. P., Von Niederhausern, A. C., Wade, C. M., Wall, M., Weber, R. J., Weiss, R. B., Wendl, M. C., West, A. P., Wetterstrand, K., Wheeler, R., Whelan, S., Wierzbowski, J., Willey, D., Williams, S., Wilson, R. K., Winter, E., Worley, K. C., Wyman, D., Yang, S., Yang, S.-P., Zdobnov, E. M., Zody, M. C., and Lander, E. S. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature*, 420(6915):520–562. (Cited on page 9.)

Mullen, R. J., Buck, C. R., and Smith, A. M. (1992). NeuN, a neuronal specific nuclear protein in vertebrates. *Development (Cambridge, England)*, 116:201–211. (Cited on page 167.)

Murray, A. J., Sauer, J.-F., Riedel, G., McClure, C., Ansel, L., Cheyne, L., Bartos, M., Wisden, W., and Wulff, P. (2011). Parvalbumin-positive CA1 interneurons are required for spatial working but not for reference memory. *Nature Neuroscience*, 14(3):297–299. (Cited on pages 271, 273, and 287.)

Myronenko, A. and Song, X. (2010). Intensity-Based Image Registration by Minimizing Residual Complexity. *IEEE Transactions on Medical Imaging*, 29(11):1882–1891. (Cited on pages 63, 66, and 111.)

Nagoshi, E., Sugino, K., Kula, E., Okazaki, E., Tachibana, T., Nelson, S., and Rosbash, M. (2010). Dissecting differential gene expression within the circadian neuronal circuit of Drosophila. *Nature Neuroscience*, 13(1):60–68. (Cited on page 8.)

Nakashiba, T., Cushman, J. D., Pelkey, K. A., Renaudineau, S., Buhl, D. L., McHugh, T. J., Barrera, V. R., Chittajallu, R., Iwamoto, K. S., McBain, C. J., Fanselow, M. S., and Tonegawa, S. (2012). Young Dentate Granule Cells Mediate Pattern Separation, whereas Old Granule Cells Facilitate Pattern Completion. *Cell*. (Cited on pages 100 and 293.)

Nakashiba, T., Young, J. Z., McHugh, T. J., Buhl, D. L., and Tonegawa, S. (2008). Transgenic inhibition of synaptic transmission reveals role of CA3 output in hippocampal learning. *Science (New York, NY)*, 319(5867):1260–1264. (Cited on pages 20 and 267.)

Narayanan, R. and Johnston, D. (2008). The ascent of channels with memory. *Neuron*, 60(5):735–738. (Cited on page 9.)

Nedivi, E., Hevroni, D., Naot, D., Israeli, D., and Citri, Y. (1993). Numerous candidate plasticity-related genes revealed by differential cDNA cloning. *Nature*, 363(6431):718–722. (Cited on page 32.)

Ng, L., Bernard, A., Lau, C., Overly, C. C., Dong, H.-W., Kuan, C., Pathak, S., Sunkin, S. M., Dang, C., Bohland, J. W., Bokil, H., Mitra, P. P., Puelles, L., Hohmann, J., Anderson, D. J., Lein, E. S., Jones, A. R., and Hawrylycz, M. (2009). An anatomic gene expression atlas of the adult mouse brain. *Nature Neuroscience*, 12(3):356–362. (Cited on pages 5, 48, and 54.)

Ng, L., Lau, C., Young, R., Pathak, S., and Kuan, L. (2007a). NeuroBlast: a 3D spatial homology search tool for gene expression. *BMC Bioinformatics*. (Cited on page 5.)

Ng, L., Pathak, S. D., Kuan, C., Lau, C., Dong, H., Sodt, A., Dang, C., Avants, B., Yushkevich, P., Gee, J. C., Haynor, D., Lein, E., Jones, A., and Hawrylycz, M. (2007b). Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain. *IEEE/ACM transactions on computational biology and bioinformatics / IEEE, ACM*, 4(3):382–393. (Cited on pages 46 and 57.)

Niere, F., Wilkerson, J. R., and Huber, K. M. (2012). Evidence for a fragile X mental retardation protein-mediated translational switch in metabotropic glutamate receptor-triggered Arc translation and long-term depression. *The Journal of Neuroscience*, 32(17):5924–5936. (Cited on pages 38, 159, and 260.)

Niewoehner, B., Single, F. N., Hvalby, Ø., Jensen, V., Meyer Zum Alten Borgloh, S., Seeburg, P. H., Rawlins, J. N. P., Sprengel, R., and Bannerman, D. M. (2007). Impaired spatial working memory but spared spatial reference memory following functional loss of NMDA receptors in the dentate gyrus. *The European Journal of Neuroscience*, 25(3):837–846. (Cited on pages 14 and 16.)

Nolan, M. F., Dudman, J. T., Dodson, P. D., and Santoro, B. (2007). HCN1 channels control resting and active integrative properties of stellate cells from layer II of the entorhinal cortex. *The Journal of Neuroscience*, 27(46):12440–12451. (Cited on page 22.)

O'Donnell, C. and Nolan, M. F. (2011). Tuning of synaptic responses: an organizing principle for optimization of neural circuits. *Trends in Neurosciences*, 34(2):51–60. (Cited on page 9.)

O'Keefe, J. (1976). Place units in the hippocampus of the freely moving rat. *Experimental neurology*, 51(1):78–109. (Cited on pages 15, 254, and 256.)

O'Keefe, J. and Dostrovsky, J. (1971). The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat. *Brain Research*, 34(1):171–175. (Cited on pages 1, 14, and 266.)

Okuno, H., Akashi, K., Ishii, Y., Yagishita-Kyo, N., Suzuki, K., Nonaka, M., Kawashima, T., Fujii, H., Takemoto-Kimura, S., Abe, M., Natsume, R., Chowdhury, S., Sakimura, K., Worley, P. F., and Bito, H. (2012). Inverse synaptic tagging of inactive synapses via dynamic interaction of Arc/Arg3.1 with CaMKIIβ. *Cell*, 149(4):886–898. (Cited on page 39.)

Olson, E. N., Arnold, H. H., Rigby, P. W., and Wold, B. J. (1996). Know your neighbors: three phenotypes in null mutants of the myogenic bHLH gene MRF4. *Cell*, 85(1):1–4. (Cited on

pages 212 and 260.)

Ons, S., Martí, O., and Armario, A. (2004). Stress-induced activation of the immediate early gene Arc (activity-regulated cytoskeleton-associated protein) is restricted to telencephalic areas in the rat brain: relationship to c-fos mRNA. *Journal of neurochemistry*, 89(5):1111–1118. (Cited on pages 33, 34, 35, 158, 215, 255, and 257.)

Panja, D., Dagyte, G., Bidinosti, M., Wibrand, K., Kristiansen, A.-M., Sonenberg, N., and Bramham, C. R. (2009). Novel translational control in Arc-dependent long term potentiation consolidation in vivo. *The Journal of biological chemistry*, 284(46):31498–31511. (Cited on pages 37 and 38.)

Pantazopoulos, H., Woo, T.-U. W., Lim, M. P., Lange, N., and Berretta, S. (2010). Extracellular matrix-glial abnormalities in the amygdala and entorhinal cortex of subjects diagnosed with schizophrenia. *Archives of General Psychiatry*, 67(2):155–166. (Cited on page 30.)

Park, S., Park, J. M., Kim, S., Kim, J.-A., Shepherd, J. D., Smith-Hicks, C. L., Chowdhury, S., Kaufmann, W., Kuhl, D., Ryazanov, A. G., Huganir, R. L., Linden, D. J., and Worley, P. F. (2008). Elongation factor 2 and fragile X mental retardation protein control the dynamic translation of Arc/Arg3.1 essential for mGluR-LTD. *Neuron*, 59(1):70–83. (Cited on page 38.)

Pastoll, H., Ramsden, H. L., and Nolan, M. F. (2012). Intrinsic electrophysiological properties of entorhinal cortex stellate cells and their contribution to grid cell firing fields. *Frontiers in neural circuits*, 6:17. (Cited on pages 21 and 22.)

Pastoll, H., Solanka, L., van Rossum, M. C. W., and Nolan, M. F. (2013). Feedback inhibition enables θ-nested γ oscillations and grid firing fields. *Neuron*, 77(1):141–154. (Cited on pages 22, 26, and 108.)

Paxinos, G. and Franklin, K. B. J. (2001). *The mouse brain in stereotaxic coordinates / George Paxinos, Keith B.J. Franklin*. Academic, San Diego, Calif. ; London :, 2nd ed. edition. (Cited on page 58.)

Peebles, C. L., Yoo, J., Thwin, M. T., Palop, J. J., Noebels, J. L., and Finkbeiner, S. (2010). Arc regulates spine morphology and maintains network stability in vivo. *Proceedings of the National Academy of Sciences of the United States of America*, 107(42):18173–18178. (Cited on pages 39, 211, and 260.)

Pikkarainen, M., Rönkkö, S., Savander, V., Insausti, R., and Pitkänen, A. (1999). Projections from the lateral, basal, and accessory basal nuclei of the amygdala to the hippocampal formation in rat. *The Journal of Comparative Neurology*, 403(2):229–260. (Cited on page 19.)

Pinaud, R., Penner, M. R., Robertson, H. A., and Currie, R. W. (2001). Upregulation of the immediate early gene arc in the brains of rats exposed to environmental enrichment: implications for molecular plasticity. *Molecular Brain Research*, 91(1-2):50–56. (Cited on page 34.)

Pintchovski, S. A., Peebles, C. L., Kim, H. J., Verdin, E., and Finkbeiner, S. (2009). The serum response factor and a putative novel transcription factor regulate expression of the immediate-early gene Arc/Arg3.1 in neurons. *The Journal of Neuroscience*, 29(5):1525–1537. (Cited on pages 33, 36, and 37.)

Plath, N., Ohana, O., Dammermann, B., Errington, M. L., Schmitz, D., Gross, C., Mao, X., Engelsberg, A., Mahlke, C., and Welzl, H. (2006). Arc/Arg3.1 Is Essential for the Consolidation of Synaptic Plasticity and Memories. *Neuron*, 52(3):437–444. (Cited on pages 38, 39, 158, 209, 213, 215, 254, and 261.)

Quinn, B., Toga, A. W., Motamed, S., and Merlic, C. A. (1995). Fluoro nissl green: a novel fluorescent counterstain for neuroanatomy. *Neuroscience letters*, 184(3):169–172. (Cited on page 212.)

R Development Core Team (2008). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. (Cited on page 186.)

Ramanan, N., Shen, Y., Sarsfield, S., Lemberger, T., Schütz, G., Linden, D. J., and Ginty, D. D. (2005). SRF mediates activity-induced gene expression and synaptic plasticity but not neuronal viability. *Nature Neuroscience*, 8(6):759–767. (Cited on page 37.)

Ramirez, J. J., Poulton, W. E., Knelson, E., Barton, C., King, M. A., and Klein, R. L. (2011). Focal expression of mutated tau in entorhinal cortex neurons of rats impairs spatial working memory. *Behavioural brain research*, 216(1):332–340. (Cited on page 30.)

Rao, V. R., Pintchovski, S. A., Chin, J., Peebles, C. L., Mitra, S., and Finkbeiner, S. (2006). AMPA receptors regulate transcription of the plasticity-related immediate-early gene Arc. *Nature Neuroscience*, 9(7):887–895. (Cited on pages 37 and 211.)

Rapanelli, M., Lew, S. E., Frick, L. R., and Zanutto, B. S. (2010). Plasticity in the rat prefrontal cortex: linking gene expression and an operant learning with a computational theory. *PloS one*, 5(1):e8656. (Cited on pages 10 and 32.)

Remondes, M. and Schuman, E. M. (2004). Role for a cortical input to hippocampal area CA1 in the consolidation of a long-term memory. *Nature*, 431(7009):699–703. (Cited on pages 16, 23, and 291.)

Rial Verde, E. M., Lee-Osbourne, J., Worley, P. F., Malinow, R., and Cline, H. T. (2006). Increased Expression of the Immediate-Early Gene Arc/Arg3.1 Reduces AMPA Receptor-Mediated Synaptic Transmission. *Neuron*, 52(3):461–474. (Cited on page 38.)

Rodríguez, J. J., Davies, H. A., Silva, A. T., De Souza, I. E. J., Peddie, C. J., Colyer, F. M., Lancashire, C. L., Fine, A., Errington, M. L., Bliss, T. V. P., and Stewart, M. G. (2005). Long-term potentiation in the rat dentate gyrus is associated with enhanced Arc/Arg3.1 protein expression in spines, dendrites and glia. *The European Journal of Neuroscience*, 21(9):2384–2396. (Cited on pages 33 and 219.)

Rowland, D. C., Weible, A. P., Wickersham, I. R., Wu, H.-Y., Seung, H. S., Mayford, M., Witter, M. P., and Kentros, C. G. (2011). Quantitative mapping of monosynaptic inputs to entorhinal layer II neurons via transgenically-targeted rabies virus suggests a strong direct projection from hippocampal area CA2. In *Society for Neuroscience*, Washington, DC. (Cited on pages 15 and 290.)

Royer, S., Sirota, A., Patel, J., and Buzsáki, G. (2010). Distinct representations and theta dynamics in dorsal and ventral hippocampus. *The Journal of Neuroscience*, 30(5):1777–1787. (Cited on page 28.)

Safran, M., Dalah, I., Alexander, J., Rosen, N., Stein, T. I., Shmoish, M., Nativ, N., Bahir, I., Doniger, T., Krug, H., Sirota-Madi, A., Olender, T., Golan, Y., Stelzer, G., Harel, A., and Lancet, D. (2010). GeneCards Version 3: the human gene integrator. *Database*, 2010. (Cited on page 115.)

Saha, R. N., Wissink, E. M., Bailey, E. R., Zhao, M., Fargo, D. C., Hwang, J.-Y., Daigle, K. R., Fenn, J. D., Adelman, K., and Dudek, S. M. (2011). Rapid activity-induced transcription of Arc and other IEGs relies on poised RNA polymerase II. *Nature Neuroscience*, 14(7):848–856. (Cited on pages 114, 132, and 133.)

Sargolini, F., Fyhn, M., Hafting, T., McNaughton, B. L., Witter, M. P., Moser, M.-B., and Moser, E. I. (2006). Conjunctive representation of position, direction, and velocity in entorhinal cortex. *Science (New York, NY)*, 312(5774):758–762. (Cited on pages 17, 19, 20, 24, 216, 258, 291, and 293.)

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature methods*, 9(7):676–682. (Cited on pages 59, 63, and 173.)

Schmidt, E. F., Kus, L., Gong, S., and Heintz, N. (2013). BAC transgenic mice and the GENSAT database of engineered mouse strains. *Cold Spring Harbor protocols*, 2013(3). (Cited on page 5.)

Schmidt-Hieber, C. and Häusser, M. (2013). Cellular mechanisms of spatial navigation in the medial entorhinal cortex. *Nature Neuroscience*, 16(3):325–331. (Cited on page 262.)

Schon, E. A. and Area-Gomez, E. (2013). Mitochondria-associated ER membranes in Alzheimer disease. *Molecular and cellular neurosciences*, 55:26–36. (Cited on page 154.)

Schwartz, S. P. and Coleman, P. D. (1981). Neurons of origin of the perforant path. *Experimental neurology*, 74(1):305–312. (Cited on pages 21 and 209.)

Schwerdtfeger, W. K., Buhl, E. H., and Germroth, P. (1990). Disynaptic olfactory input to the hippocampus mediated by stellate cells in the entorhinal cortex. *The Journal of Comparative Neurology*, 292(2):163–177. (Cited on page 6.)

Scoville, W. B. and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of neurology, neurosurgery, and psychiatry*, 20(1):11–21. (Cited on pages 13 and 14.)

Shah, M. M., Anderson, A. E., Leung, V., Lin, X., and Johnston, D. (2004). Seizure-Induced Plasticity of h Channels in Entorhinal Cortical Layer III Pyramidal Neurons. *Neuron*, 44(3):495–508. (Cited on page 31.)

Sharp, P. E. and Green, C. (1994). Spatial correlates of firing patterns of single cells in the subiculum of the freely moving rat. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 14(4):2339–2356. (Cited on page 14.)

Shen, Y., Yue, F., McCleary, D. F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature*, 488(7409):116–120. (Cited on page 10.)

Shepherd, J. D., Rumbaugh, G., Wu, J., Chowdhury, S., Plath, N., Kuhl, D., Huganir, R. L., and Worley, P. F. (2006). Arc/Arg3.1 mediates homeostatic synaptic scaling of AMPA receptors. *Neuron*, 52(3):475–484. (Cited on pages 38, 158, and 215.)

Siegert, S., Cabuy, E., Scherf, B. G., Kohler, H., Panda, S., Le, Y.-Z., Fehling, H. J., Gaidatzis, D., Stadler, M. B., and Roska, B. (2012). Transcriptional code and disease map for adult retinal cell types. *Nature Neuroscience*. (Cited on pages 8 and 291.)

Smith-Hicks, C., Xiao, B., Deng, R., Ji, Y., Zhao, X., Shepherd, J. D., Posern, G., Kuhl, D., Huganir, R. L., Ginty, D. D., Worley, P. F., and Linden, D. J. (2010). SRF binding to SRE 6.9 in the Arc promoter is essential for LTD in cultured Purkinje cells. *Nature Neuroscience*, 13(9):1082–1089. (Cited on pages 37 and 38.)

Solodkin, A. and Van Hoesen, G. W. (1996). Entorhinal cortex modules of the human brain. *The Journal of Comparative Neurology*, 365(4):610–617. (Cited on pages 19 and 20.)

Solodkin, A., Veldhuizen, S. D., and Van Hoesen, G. W. (1996). Contingent vulnerability of entorhinal parvalbumin-containing neurons in Alzheimer's disease. *The Journal of Neuroscience*, 16(10):3311–3321. (Cited on pages 153 and 154.)

Solstad, T., Boccara, C. N., Kropff, E., Moser, M.-B., and Moser, E. I. (2008). Representation of geometric borders in the entorhinal cortex. *Science (New York, NY)*, 322(5909):1865–1868. (Cited on page 17.)

Steffenach, H.-A., Witter, M., Moser, M.-B., and Moser, E. I. (2005). Spatial memory in the rat requires the dorsolateral band of the entorhinal cortex. *Neuron*, 45(2):301–313. (Cited on pages 14 and 28.)

Stensola, H., Stensola, T., Solstad, T., Frøland, K., Moser, M.-B., and Moser, E. I. (2012). The entorhinal grid map is discretized. *Nature*, 492(7427):72–78. (Cited on pages 17, 26, 109, 129, 138, 151, 158, 208, 290, 292, and 294.)

Sternberg (1983). Biomedical Image Processing. *Computer*, 16(1):22–34. (Cited on page 174.)

Steward, O. and Scoville, S. A. (1976). Cells of origin of entorhinal cortical afferents to the hippocampus and fascia dentata of the rat. *The Journal of Comparative Neurology*, 169(3):347–370. (Cited on pages 15 and 108.)

Steward, O. and Worley, P. F. (2001a). A cellular mechanism for targeting newly synthesized mRNAs to synaptic sites on dendrites. *Proceedings of the National Academy of Sciences of the United States of America*, 98(13):7062–7068. (Cited on pages 33 and 38.)

Steward, O. and Worley, P. F. (2001b). Selective targeting of newly synthesized Arc mRNA to active synapses requires NMDA receptor activation. *Neuron*, 30(1):227–240. (Cited on pages 32 and 37.)

Strittmatter, W. J., Saunders, A. M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G. S., and Roses, A. D. (1993). Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. *Proceedings of the National Academy of Sciences of the United States of America*, 90(5):1977–1981. (Cited on page 11.)

Sugar, J., Witter, M. P., van Strien, N. M., and Cappaert, N. L. M. (2011). The retrosplenial cortex: intrinsic connectivity and connections with the (para)hippocampal region in the rat. An interactive connectome. *Frontiers in Neuroinformatics*, 5:7. (Cited on page 24.)

Sugino, K., Hempel, C. M., Miller, M. N., Hattox, A. M., Shapiro, P., Wu, C., Huang, Z. J., and Nelson, S. B. (2006). Molecular taxonomy of major neuronal classes in the adult mouse forebrain. *Nature Neuroscience*, 9(1):99–107. (Cited on pages 4, 6, 7, 8, and 291.)

Suthana, N., Haneef, Z., Stern, J., Mukamel, R., Behnke, E., Knowlton, B., and Fried, I. (2012). Memory Enhancement and Deep-Brain Stimulation of the Entorhinal Area. *New England Journal of Medicine*, 366(6):502–510. (Cited on pages 16 and 108.)

Sutherland, R. J., Whishaw, I. Q., and Kolb, B. (1988). Contributions of cingulate cortex to two forms of spatial learning and memory. *The Journal of Neuroscience*, 8(6):1863–1872. (Cited on page 25.)

Suzuki, W. A. and Porteros, A. (2002). Distribution of calbindin D-28k in the entorhinal, perirhinal, and parahippocampal cortices of the macaque monkey. *The Journal of Comparative Neurology*, 451(4):392–412. (Cited on page 19.)

Swanson, L. W. and Cowan, W. M. (1977). An autoradiographic study of the organization of the efferent connections of the hippocampal formation in the rat. *The Journal of Comparative Neurology*, 172(1):49–84. (Cited on pages 15, 24, 100, 108, 153, 209, 258, 266, 286, 289, and 294.)

Swanson, L. W., Wyss, J. M., and Cowan, W. M. (1978). An autoradiographic study of the organization of intrahippocampal association pathways in the rat. *The Journal of Comparative Neurology*, 181(4):681–715. (Cited on pages 230 and 274.)

Takehara-Nishiuchi, K., Insel, N., Hoang, L. T., Wagner, Z., Olson, K., Chawla, M. K., Burke, S. N., and Barnes, C. A. (2012). Activation Patterns in Superficial Layers of Neocortex Change Between Experiences Independent of Behavior, Environment, or the Hippocampus. *Cerebral Cortex*. (Cited on pages 216 and 288.)

Tamamaki, N. and Nojyo, Y. (1993). Projection of the entorhinal layer II neurons in the rat as revealed by intracellular pressure-injection of neurobiotin. *Hippocampus*, 3(4):471–480. (Cited on page 36.)

Tamamaki, N. and Nojyo, Y. (1995). Preservation of topography in the connections between the subiculum, field CA1, and the entorhinal cortex in rats. *The Journal of Comparative Neurology*, 353(3):379–390. (Cited on pages 15, 24, 100, 108, 153, 216, 266, 267, and 294.)

Tanzi, R. E. and Bertram, L. (2005). Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell*, 120(4):545–555. (Cited on page 30.)

Taube, J. S. (1995). Place cells recorded in the parasubiculum of freely moving rats. *Hippocampus*, 5(6):569–583. (Cited on page 14.)

Taube, J. S. (2007). The head direction signal: origins and sensory-motor integration. *Annual review of neuroscience*, 30:181–207. (Cited on page 19.)

Thompson, C. L., Pathak, S. D., Jeromin, A., Ng, L. L., MacPherson, C. R., Mortrud, M. T., Cusick, A., Riley, Z. L., Sunkin, S. M., Bernard, A., Puchalski, R. B., Gage, F. H., Jones, A. R., Bajic, V. B., Hawrylycz, M. J., and Lein, E. S. (2008). Genomic anatomy of the hippocampus. *Neuron*, 60(6):1010–1021. (Cited on pages 9, 43, 49, 100, 142, and 143.)

Thompson, C. L., Wisor, J. P., Lee, C.-K., Pathak, S. D., Gerashchenko, D., Smith, K. A., Fischer, S. R., Kuan, C. L., Sunkin, S. M., Ng, L. L., Lau, C., Hawrylycz, M., Jones, A. R., Kilduff, T. S., and Lein, E. S. (2010). Molecular and anatomical signatures of sleep deprivation in the mouse brain. *Frontiers in neuroscience*, 4:165. (Cited on pages 35, 208, and 215.)

Toledo-Rodriguez, M., Goodman, P., Illic, M., Wu, C., and Markram, H. (2005). Neuropeptide and calcium-binding protein gene expression profiles predict neuronal anatomical type in the juvenile rat. *The Journal of Physiology*, 567(Pt 2):401–413. (Cited on page 8.)

Tolner, E. A., Kloosterman, F., Van Vliet, E. A., Witter, M. P., Silva, F. H. L. d., and Gorter, J. A. (2005). Presubiculum stimulation in vivo evokes distinct oscillations in superficial and deep entorhinal cortex layers in chronic epileptic rats. *The Journal of Neuroscience*, 25(38):8755–8765. (Cited on pages 23 and 31.)

Tsao, A., Moser, M.-B., and Moser, E. I. (2013). Traces of experience in the lateral entorhinal cortex. *Current biology : CB*, 23(5):399–405. (Cited on page 13.)

Tu, B., Gu, Z., Shen, J.-x., Lamb, P. W., and Yakel, J. L. (2009). Characterization of a nicotine-sensitive neuronal population in rat entorhinal cortex. *The Journal of Neuroscience*, 29(33):10436–10448. (Cited on page 109.)

Tzingounis, A. V. and Nicoll, R. A. (2006). Arc/Arg3.1: linking gene expression to synaptic plasticity and memory. *Neuron*, 52(3):403–407. (Cited on pages 33 and 209.)

van Groen, T. (2001). Entorhinal cortex of the mouse: cytoarchitectonical organization. *Hippocampus*, 11(4):397–407. (Cited on page 42.)

van Groen, T., Miettinen, P., and Kadish, I. (2003). The entorhinal cortex of the mouse: organization of the projection to the hippocampal formation. *Hippocampus*, 13(1):133–149. (Cited on pages 6, 28, 29, 100, 108, 209, 266, and 289.)

van Haeften, T., Baks-te Bulte, L., Goede, P. H., Wouterlood, F. G., and Witter, M. P. (2003). Morphological and numerical analysis of synaptic interactions between neurons in deep and superficial layers of the entorhinal cortex of the rat. *Hippocampus*, 13(8):943–952. (Cited on pages 6 and 19.)

van Haeften, T., Wouterlood, F. G., Jorritsma-Byham, B., and Witter, M. P. (1997). GABAergic Presubicular Projections to the Medial Entorhinal Cortex of the Rat. *The Journal of Neuroscience*. (Cited on page 23.)

van Praag, H., Schinder, A. F., Christie, B. R., Toni, N., Palmer, T. D., and Gage, F. H. (2002). Functional neurogenesis in the adult hippocampus. *Nature*, 415(6875):1030–1034. (Cited on page 108.)

van Strien, N. M., Cappaert, N. L. M., and Witter, M. P. (2009). The anatomy of memory: an interactive overview of the parahippocampal-hippocampal network. *Nature Reviews Neuroscience*, 10(4):272–282. (Cited on pages 13, 14, 16, and 28.)

Varga, C., Lee, S. Y., and Soltesz, I. (2010). Target-selective GABAergic control of entorhinal cortex output. *Nature Neuroscience*, 13(7):822–824. (Cited on pages 21, 22, 23, 106, 108, 137, and 152.)

Vedaldi, A. and Fulkerson, B. (2008). VLFeat: An Open and Portable Library of Computer Vision Algorithms . (Cited on page 61.)

Vedaldi, A. and Zisserman, A. (2013). Practical Image Classification. (Cited on page 61.)

von Hertzen, L. S. J. and Giese, K. P. (2005). Memory reconsolidation engages only a subset of immediate-early genes induced during consolidation. *The Journal of Neuroscience*, 25(8):1935–1942. (Cited on page 32.)

Wallace, C. S., Lyford, G. L., Worley, P. F., and Steward, O. (1998). Differential intracellular sorting of immediate early gene mRNAs depends on signals in the mRNA sequence. *The Journal of Neuroscience*, 18(1):26–35. (Cited on pages 219 and 261.)

Walling, S. G., Bromley, K., and Harley, C. W. (2006). Glycogen phosphorylase reactivity in the entorhinal complex in familiar and novel environments: evidence for labile glycogenolytic modules in the rat. *Journal of chemical neuroanatomy*, 31(2):108–113. (Cited on page 21.)

Wang, K. H., Majewska, A., Schummers, J., Farley, B., Hu, C., Sur, M., and Tonegawa, S. (2006). In vivo two-photon imaging reveals a role of arc in enhancing orientation specificity in visual cortex. *Cell*, 126(2):389–402. (Cited on pages 35, 39, 158, 159, 160, 161, 163, 180, 181, 182, 185, 208, 210, 212, 216, 217, 218, 219, 255, 259, 260, 261, and 292.)

Wang, Y., Zheng, F., Zhou, X., Sun, Z., and Wang, H. (2009a). Converging signal on ERK1/2 activity regulates group I mGluR-mediated Arc transcription. *Neuroscience letters*, 460(1):36–40. (Cited on page 37.)

Wang, Z., Gerstein, M., and Snyder, M. (2009b). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics*, 10(1):57–63. (Cited on page 5.)

Waung, M. W., Pfeiffer, B. E., Nosyreva, E. D., Ronesi, J. A., and Huber, K. M. (2008). Rapid translation of Arc/Arg3.1 selectively mediates mGluR-dependent LTD through persistent increases in AMPAR endocytosis rate. *Neuron*, 59(1):84–97. (Cited on pages 38, 158, 212, and 216.)

Weible, A. P., Schwarcz, L., Wickersham, I. R., Deblander, L., Wu, H., Callaway, E. M., Seung, H. S., and Kentros, C. G. (2010). Transgenic targeting of recombinant rabies virus reveals monosynaptic connectivity of specific neurons. *The Journal of Neuroscience*, 30(49):16509–16513. (Cited on page 15.)

White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode Caenorhabditis elegans. *Philosophical transactions of the Royal Society of London Series B, Biological sciences*, 314(1165):1–340. (Cited on page 6.)

Wilson, M. A. and McNaughton, B. L. (1993). Dynamics of the hippocampal ensemble code for space. *Science (New York, NY)*, 261(5124):1055–1058. (Cited on page 35.)

Witter, M. P. (1993). Organization of the entorhinal-hippocampal system: a review of current anatomical data. *Hippocampus*, 3 Spec No:33–44. (Cited on page 15.)

Witter, M. P. (2007). Intrinsic and extrinsic wiring of CA3: indications for connectional heterogeneity. *Learning & memory (Cold Spring Harbor, N.Y.)*, 14(11):705–713. (Cited on page 289.)

Witter, M. P., Groenewegen, H. J., Lopes da Silva, F. H., and Lohman, A. H. (1989). Functional organization of the extrinsic and intrinsic circuitry of the parahippocampal region. *Progress in Neurobiology*, 33(3):161–253. (Cited on pages 15 and 19.)

Witter, M. P. and Moser, E. I. (2006). Spatial representation and the architecture of the entorhinal cortex. *Trends in Neurosciences*, 29:671–678. (Cited on pages 13 and 19.)

Wood, E. R., Dudchenko, P. A., and Eichenbaum, H. (1999). The global record of memory in hippocampal neuronal activity. *Nature*, 397(6720):613–616. (Cited on page 14.)

Yamagata, K., Andreasson, K. I., Kaufmann, W. E., Barnes, C. A., and Worley, P. F. (1993). Expression of a mitogen-inducible cyclooxygenase in brain neurons: regulation by synaptic activity and glucocorticoids. *Neuron*, 11(2):371–386. (Cited on page 32.)

Yartsev, M. M., Witter, M. P., and Ulanovsky, N. (2011). Grid cells without theta oscillations in the entorhinal cortex of bats. *Nature*, 479(7371):103–107. (Cited on page 17.)

Zeng, H., Shen, E. H., Hohmann, J. G., Oh, S. W., Bernard, A., Royall, J. J., Glattfelder, K. J., Sunkin, S. M., Morris, J. A., Guillozet-Bongaarts, A. L., Smith, K. A., Ebbert, A. J., Swanson, B., Kuan, L., Page, D. T., Overly, C. C., Lein, E. S., Hawrylycz, M. J., Hof, P. R., Hyde, T. M., Kleinman, J. E., and Jones, A. R. (2012). Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. *Cell*, 149(2):483–496. (Cited on page 7.)

Zhang, S.-J., Ye, J., Miao, C., Tsao, A., Cerniauskas, I., Ledergerber, D., Moser, M.-B., and Moser, E. I. (2013). Optogenetic dissection of entorhinal-hippocampal functional connectivity. *Science (New York, NY)*, 340(6128):1232627. (Cited on pages 23 and 293.)

Zhang, Z., Reboreda, A., Alonso, A., Barker, P. A., and Séguéla, P. (2011). TRPC channels underlie cholinergic plateau potentials and persistent activity in entorhinal cortex. *Hippocampus*, 21(4):386–397. (Cited on page 109.)

Ziv, Y., Burns, L. D., Cocker, E. D., Hamel, E. O., Ghosh, K. K., Kitch, L. J., El Gamal, A., and Schnitzer, M. J. (2013). Long-term dynamics of CA1 hippocampal place codes. *Nature Neuroscience*, 16(3):264–266. (Cited on page 14.)

COLOPHON

This document was typeset using the typographical look-and-feel `classicthesis` developed by André Miede. The style was inspired by Robert Bringhurst's seminal book on typography "*The Elements of Typographic Style*". `classicthesis` is available for both LaTeX and LyX:

http://code.google.com/p/classicthesis/

Happy users of `classicthesis` usually send a real postcard to the author, a collection of postcards received so far is featured here:

http://postcards.miede.de/

Final Version as of March 10, 2014 (`classicthesis` version 4.0).